CUK-H03676-11-020212U9

ANNUAL REVIEW OF MEDICINE:

Selected Topics in the Clinical Sciences

11

EDITORIAL COMMITTEE (1981)

C. H. COGGINS
Harvard Medical School
W. P. CREGER
Stanford University
R. M. DONALDSON, JR.
Yale University
E. W. HANCOCK
Stanford University
P. G. QUIE
University of Minnesota
R. M. ROSE
University of Texas
P. S. RUSSELL
Harvard Medical School
J. D. WILSON
University of Texas

Responsible for the organization of Volume 32
(Editorial Committee, 1979)

C. H. COGGINS
Harvard Medical School
W. P. CREGER
Stanford University
R. M. DONALDSON, JR.
Yale University
E. W. HANCOCK
Stanford University
P. G. QUIE
University of Minnesota
R. M. ROSE
University of Texas
M. SIPERSTEIN
University of California Service
Veterans Administration Hospital,
San Francisco
J. A. COLLINS (Guest)
Stanford University
R. JONES (Guest)
University of California
San Francisco

Production Editor M. E. PLATT

ANNUAL REVIEW OF MEDICINE:

Selected Topics in the Clinical Sciences

WILLIAM P. CREGER, *Editor*
Stanford University School of Medicine

C. H. COGGINS, *Associate Editor*
Harvard Medical School

E. W. HANCOCK, *Associate Editor*
Stanford University School of Medicine

VOLUME 32

1981

ANNUAL REVIEWS INC. 4139 EL CAMINO WAY PALO ALTO, CALIFORNIA 94306 USA

ANNUAL REVIEWS INC.
Palo Alto, California, USA

COPYRIGHT © 1981 BY ANNUAL REVIEWS INC., PALO ALTO, CALIFORNIA, USA. ALL RIGHTS RESERVED. The appearance of the code at the bottom of the first page of an article in this serial indicates the copyright owner's consent that copies of the article may be made for personal or internal use, or for the personal or internal use of specific clients. This consent is given on the condition, however, that the copier pay the stated per-copy fee of $1.00 per article through the Copyright Clearance Center, Inc. (21 Congress Street, Salem, MA 01970) for copying beyond that permitted by Sections 107 or 108 of the US Copyright Law. The per-copy fee of $1.00 per article also applies to the copying, under the stated conditions, of articles published in any Annual Review serial before January 1, 1978. Individual readers, and nonprofit libraries acting for them, are permitted to make a single copy of an article without charge for use in research or teaching. This consent does not extend to other kinds of copying, such as copying for general distribution, for advertising or promotional purposes, for creating new collective works, or for resale. For such uses, written permission is required. Write to Permissions Dept., Annual Reviews Inc., 4139 El Camino Way, Palo Alto, CA 94306 USA.

REPRINTS The conspicuous number aligned in the margin with the title of each article in this volume is a key for use in ordering reprints. Available reprints are priced at the uniform rate of $2.00 each postpaid. The minimum acceptable reprint order is 5 reprints and/or $10.00 prepaid. A quantity discount is available.

International Standard Serial Number: 0066-4219
International Standard Book Number: 0-8243-0532-9
Library of Congress Catalog Card Number: A51-1659

Gx 21249

Annual Reviews Inc. and the Editors of its publications assume no responsibility for the statements expressed by the contributors to this Review.

PRINTED AND BOUND IN THE UNITED STATES OF AMERICA

Annual Review of Medicine
Volume 32, 1981

CONTENTS

[Note: Titles of chapters in Volumes 28–32 are arranged by category on pages 618–24.]

SEROLOGICAL MARKERS OF HEPATITIS B VIRUS INFECTION, *Jay H. Hoofnagle* 1
IMMUNE COMPLEXES IN HUMAN DISEASES, *Ralph C. Williams, Jr.* 13
BACTERIAL SURFACE COMPONENTS AND THE PATHOGENESIS OF INFECTIOUS DISEASES, *Phillip K. Peterson and Paul G. Quie* 29
CHLAMYDIAL INFECTIONS, *J. Schachter and M. Grossman* 45
CHEMICAL HEADACHES, *Neil H. Raskin* 63
PAPILLARY AND FOLLICULAR THYROID CANCER: A Selective Approach to Diagnosis and Treatment, *Ernest L. Mazzaferri* 73
GENETIC COMPONENT OF ALCOHOLISM, *Donald W. Goodwin* 93
RETINAL VASCULAR OCCLUSION, *Walter H. Stern and Desmond B. Archer* 101
HOST DETERMINANTS IN STAPHYLOCOCCAL DISEASE, *Jan Verhoef and Henri A. Verbrugh* 107
NEUROENDOCRINOLOGICAL EFFECTS OF ALCOHOL, *Theodore J. Cicero* 123
ANEMIA OF PREMATURITY, *Peter R. Dallman* 143
AMEBIC LIVER ABSCESS, *R. S. Peters, N. Gitlin, and R. D. Libke* 161
RELATIONS BETWEEN CIRCULATORY AND METABOLIC CHANGES IN SEPSIS, *John H. Siegel* 175
TREATMENT OF SICKLE CELL ANEMIA, *Samuel Charache* 195
OUTLOOK FOR ACUTE LYMPHOCYTIC LEUKEMIA IN CHILDREN IN 1982, *Joseph V. Simone* 207
CARDIAC TRANSPLANTATION, *Sharon A. Hunt and Edward B. Stinson* 213
THE PROBLEM OF OBESITY, *Faith T. Fitzgerald* 221
TOTAL PARENTERAL NUTRITION IN THE MANAGEMENT OF THE CANCER PATIENT, *Murray F. Brennan* 233

MAGNESIUM DEFICIENCY AND EXCESS, *R. K. Rude and F. R. Singer* 245
AMYLOIDOSIS, *Peter D. Gorevic and Edward C. Franklin* 261
PATHOPHYSIOLOGY OF PREECLAMPSIA, *Marshall D. Lindheimer and Adrian I. Katz* 273
PULMONARY EDEMA DUE TO INCREASED MICROVASCULAR PERMEABILITY, *Norman C. Staub* 291
THE KILLING OF PATHOGENS BY PHAGOCYTES, *T. G. Gabig and B. M. Babior* 313
CLOSTRIDIA AND ANTIMICROBIAL ENTEROCOLITIS, *Joseph Silva, Jr. and Robert Fekety* 327
DIAGNOSIS AND TREATMENT OF HERPES SIMPLEX ENCEPHALITIS, *R. Whitley* 335
PATHOGENESIS OF ACUTE BACTERIAL DIARRHEAL DISORDERS, *Ralph A. Giannella* 341
PROSTAGLANDINS AND WATER EXCRETION, *Randall Mark Zusman* 359
PROSTAGLANDINS AND THE DUCTUS ARTERIOSUS, *P. M. Olley and F. Coceani* 375
PRIMARY RADIATION THERAPY FOR BREAST CANCER, *Jay R. Harris, Martin B. Levene, and Samuel Hellman* 387
CONVULSIVE AND DRUG THERAPIES OF DEPRESSION, *Max Fink* 405
DIETARY PRECURSORS AND BRAIN NEUROTRANSMITTER FORMATION, *John D. Fernstrom* 413
DOWNWARD TREND IN CARDIOVASCULAR MORTALITY, *Thomas J. Thom and William B. Kannel* 427
PHYSIOLOGY OF CARDIOPULMONARY RESUSCITATION, *Myron L. Weisfeldt and Nisha Chandra* 435
PATHOGENESIS, DIAGNOSIS, AND MANAGEMENT OF REFLUX ESOPHAGITIS, *D. Kestenbaum and J. Behar* 443
THE REGULATION OF RENAL CALCIUM EXCRETION: An Approach to Hypercalciuria, *Andrew F. Stewart and Arthur E. Broadus* 457
THE NATURAL HISTORY AND TREATMENT OF OVARIAN CANCER, *Dan L. Longo and Robert C. Young* 475
CHRONIC MUCOCUTANEOUS CANDIDIASIS, *John M. Dwyer* 491
PROGNOSIS IN BUNDLE BRANCH BLOCK, *J. McAnulty and S. Rahimtoola* 499
MALE ERECTILE DISORDERS, *Raul C. Schiavi* 509
DISORDERS OF GLUTAMATE METABOLISM AND NEUROLOGICAL DYSFUNCTION, *Stanley B. Prusiner* 521

TESTICULAR CANCER, *Donald G. Skinner, Peter T. Scardino, and John R. Daniels* 543
CEPHALOSPORINS, *Barbara E. Murray and Robert C. Moellering* 559
DIFFERENTIAL DIAGNOSIS AND TREATMENT OF PANIC ATTACKS AND PHOBIC STATES, *Michael R. Liebowitz and Donald F. Klein* 583
CORONARY RISK FACTORS IN THE YOUNG, *Evan A. Stein, Charles J. Glueck, and John A. Morrison* 601

CUMULATIVE INDEXES

Contributing Authors, Volumes 28–32 615
Chapter Titles, Volumes 28–32 618

ANNUAL REVIEWS INC. is a nonprofit corporation established to promote the advancement of the sciences. Beginning in 1932 with the *Annual Review of Biochemistry,* the Company has pursued as its principal function the publication of high quality, reasonably priced *Annual Review* volumes. The volumes are organized by Editors and Editorial Committees who invite qualified authors to contribute critical articles reviewing significant developments within each major discipline. Annual Reviews Inc. is administered by a Board of Directors, whose members serve without compensation.

1981 Board of Directors, Annual Reviews Inc.

Dr. J. Murray Luck, Founder and Director Emeritus of Annual Reviews Inc. *Professor Emeritus of Chemistry, Stanford University*

Dr. Joshua Lederberg, President of Annual Reviews Inc. *President, The Rockefeller University*

Dr. James E. Howell, Vice President of Annual Reviews Inc. *Professor of Economics, Stanford University*

Dr. William O. Baker, *President, Bell Telephone Laboratories*

Dr. Robert W. Berliner, *Dean, Yale University School of Medicine*

Dr. Sidney D. Drell, *Deputy Director, Stanford Linear Accelerator Center*

Dr. Eugene Garfield, *President, Institute for Scientific Information*

Dr. William D. McElroy, *Chancellor, University of California, San Diego*

Dr. William F. Miller, *President, SRI International*

Dr. Colin S. Pittendrigh, *Director, Hopkins Marine Station*

Dr. Esmond E. Snell, *Professor of Microbiology and Chemistry, University of Texas, Austin*

Dr. Harriet A. Zuckerman, *Professor of Sociology, Columbia University*

Management of Annual Reviews Inc.

John S. McNeil, Chief Executive Officer and Secretary-Treasurer
William Kaufmann, Editor-in-Chief
Sharon E. Hawkes, Production Manager
Mickey G. Hamilton, Promotion Manager

Publications

Annual Reviews of Anthropology, Astronomy and Astrophysics, Biochemistry, Biophysics and Bioengineering, Earth and Planetary Sciences, Ecology and Systematics, Energy, Entomology, Fluid Mechanics, Genetics, Materials Science, Medicine, Microbiology, Neuroscience, Nuclear and Particle Science, Nutrition, Pharmacology and Toxicology, Physical Chemistry, Physiology, Phytopathology, Plant Physiology, Psychology, Public Health, and *Sociology.*

Special Publications: *History of Entomology* (1973), *The Excitement and Fascination of Science* (1965), *The Excitement and Fascination of Science, Volume Two* (1978), *Annual Reviews Reprints: Cell Membranes, 1975–1977* (published 1978), *Annual Reviews Reprints: Immunology, 1977–1979* (published 1980), and *Intelligence and Affectivity: Their Relationship During Child Development,* by Jean Piaget (1981).

For the convenience of readers, a detachable order form/envelope is bound into the back of this volume.

SOME RELATED ARTICLES IN OTHER *ANNUAL REVIEWS*

From the *Annual Review of Neuroscience,* Volume 4 (1981):

The Biology of Myasthenia Gravis, Daniel B. Drachman

Sleep and Its Disorders, Elliot D. Weitzman

From the *Annual Review of Pharmacology and Toxicology,* Volume 20 (1980):

Advances in Ocular Pharmacology, Thom J. Zimmerman, Barry Leader, and Herbert E. Kaufman

Biology of Myasthenia Gravis, Jon Lindstrom and Peter Dau

Drug Pharmacokinetics in Cardiac and Hepatic Disease, Roger L. Williams and Leslie Z. Benet

Potential Therapeutic Usefulness of Marijuana, Louis Lemberger

Drug Therapy in Renal Failure, Marcus M. Reidenberg and Dennis E. Drayer

From the *Annual Review of Physiology,* Volume 43 (1981):

Factors Affecting Changes in the Neonatal Systemic Circulation, Michael A. Heymann, Harriet S. Iwamoto, and Abraham M. Rudolph

Regulation of the Cerebral Circulation, Hermes A. Kontos

Cardiocirculatory Dynamics in the Normal and Failing Heart, Robert Zelis, Stephen F. Flaim, A. James Liedtke, and Stephen H. Nellis

Endocrine Physiology of the Placenta, Evan R. Simpson and Paul C. MacDonald

Central Regulation of Respiration by Endogenous Neurotransmitters and Neuromodulators, Frederic L. Eldridge and David E. Millhorn

From the *Annual Review of Public Health,* Volume 2 (1981):

Maternal Behavior and Perinatal Risks: Alcohol, Smoking, and Drugs, I. M. Cushner

Issues in Antenatal and Neonatal Screening and Surveillance for Hereditary and Congenital Disorders, Neil A. Holtzman, Claire O. Leonard, and Mark R. Farfel

Prevention of Dental Disease: Caries and Peridontal Disease, Max H. Schoen and James R. Freed

Ann. Rev. Med. 1981. 32:1–11

SEROLOGIC MARKERS OF HEPATITIS B VIRUS INFECTION[1]

❖7378

Jay H. Hoofnagle, M.D.

Liver Diseases Section, Digestive Diseases Branch, National Institute of
Arthritis, Metabolism, and Digestive Diseases, National Institutes of Health,
Bethesda, Maryland 20205

The number and complexity of the various serologic markers associated with type B hepatitis make the interpretation of these markers difficult. Yet there is probably no other infectious disease for which there are so many sensitive and specific assays to define the course of the infection. Properly applied, serologic markers for type B hepatitis reveal much about the stage of infection, degree of infectivity, and immune status of the patient.

HEPATITIS B VIRUS ANTIGENS

The hepatitis B virus is a 42-nanometer, double-shelled DNA virus (Figure 1). The intact virus, referred to as the Dane particle, has an outer surface component of hepatitis B surface antigen (HBsAg, Australia antigen) and an inner core component of hepatitis B core antigen (HBcAg) (2, 5, 6). HBsAg can be detected in the serum during both acute and chronic infection with this virus, largely because this antigen circulates not only as the surface component of Dane particles, but also as incomplete HBsAg particles—20-nanometer spheres and tubules. The amount of HBsAg found in the serum can be quite high—as much as 500 $\mu g/ml$—a concentration similar to that of some major normal human serum components, such as C_4 and ceruloplasmin.

The second antigen, HBcAg, is not found free in the serum, although it can be extracted from intact Dane particles by careful purification techniques. By immunofluorescence or immunoperoxidase techniques, one can

---

[1] The US Government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this paper.

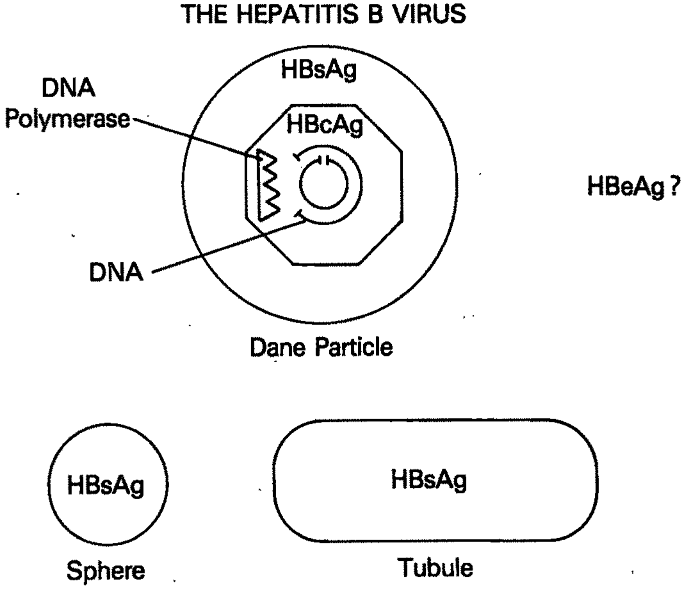


Figure 1 Diagrammatic representation of the structure of virus-like particles found in serum positive for hepatitis B surface antigen (HBsAg). The 42-nanometer (nm) Dane particle, which is the candidate hepatitis B virus, has (*a*) an outer surface component of HBsAg, and (*b*) an inner core component of hepatitis B core antigen (HBcAg), inside of which is an enzyme (DNA-polymerase) and a molecule of double-stranded, circular DNA. The 20-nm spheres and tubules are probably incomplete virus particles consisting of HBsAg alone without HBcAg or DNA. The hepatitis B e antigen (HBeAg) is a soluble hepatitis B virus antigen and is not associated with a virus-like form.

find HBcAg in the nuclei of hepatocytes from persons with hepatitis B virus infection. In contrast, HBsAg is usually found in the cytoplasm of hepatocytes (3).

The third antigen related to the hepatitis B virus is the so-called hepatitis B e antigen (HBeAg) (13). HBeAg is a soluble protein found only in HBsAg positive serum. HBeAg is not associated with a virus-like particle and its relationship to the structure of the hepatitis B virus is not yet clear; most evidence suggests that HBeAg is a subcomponent of core antigen or its breakdown product (23).

The hepatitis B virus also has an endogenous enzyme—a DNA polymerase—and a distinctive, double-stranded, circular DNA (11, 18, 21). The DNA is approximately 3200 nucleotides in length. One of the DNA strands is nicked (so that the 3' and 5' ends are not joined) and the other strand is incomplete (leaving a single-stranded segment of the DNA that averages about one third of the total length). These characteristics of the hepatitis B virus place it in a viral class of its own. Recently, similar viruses were found in woodchucks (*Marmota monax*) and ground squirrels (*Spermophilus beechyi*) (17, 22). This led to the suggestion that these agents be named "Hepadna viruses" as they share the characteristics of having a distinctive, double-stranded, circular DNA, an endogenous DNA polymerase, multiple particulate forms and antigens, and a trophism for the liver.

SEROLOGIC RESPONSES TO HEPATITIS B VIRUS INFECTION

Each of the antigens related to the hepatitis B virus can induce an antibody response in the host during acute or chronic infection. There are basically three patterns of development of these antigens and antibodies: (*a*) acute symptomatic type B hepatitis; (*b*) self-limited and subclinical hepatitis B virus infection; and (*c*) development of the chronic HBsAg carrier state (10). Why certain individuals develop acute hepatitis, while others have a subclinical infection or develop the carrier state, is not known. Host factors (age, sex, ethnic or genetic background, immunological competence) rather than viral factors (serotype or virulence) appear to be the main determinants of outcome of infection with this virus.

Acute, Symptomatic Type B Hepatitis

The serologic events that occur during symptomatic acute type B hepatitis are depicted in Figure 2. HBsAg is the first serologic marker to appear. It becomes detectable during the late incubation period of the disease, 2–5 weeks before the onset of symptoms. HBsAg rises in titer and peaks at or shortly after the onset of clinical illness. Generally, HBsAg persists for 1–5 months, declining in titer and disappearing with the improvement and resolution of clinical symptoms. Antibody to HBsAg (anti-HBs) does not become detectable during the acute disease, but rather with convalescence. Indeed, there is often a gap or "window period" between the clearance of HBsAg and the appearance of anti-HBs. Anti-HBs is a long-lived antibody, and as such, it can be used as a reliable marker of recovery and immunity to type B hepatitis.

HBcAg is not found free in the serum, but antibody to HBcAg (anti-HBc) is readily detectable. Anti-HBc appears shortly before or at the onset of clinical symptoms and quickly rises to high titers. It is the most reliable serologic marker of hepatitis B virus infection because it arises in virtually all persons infected with this virus. During the "window period" after HBsAg has disappeared, anti-HBc may be the only serologic marker of hepatitis B virus infection that is present. Initially, the anti-HBc consists of both IgG and IgM antibody. The IgM anti-HBc falls off in titer and disappears within 6–12 months of recovery (9). Anti-HBc is invaluable as a

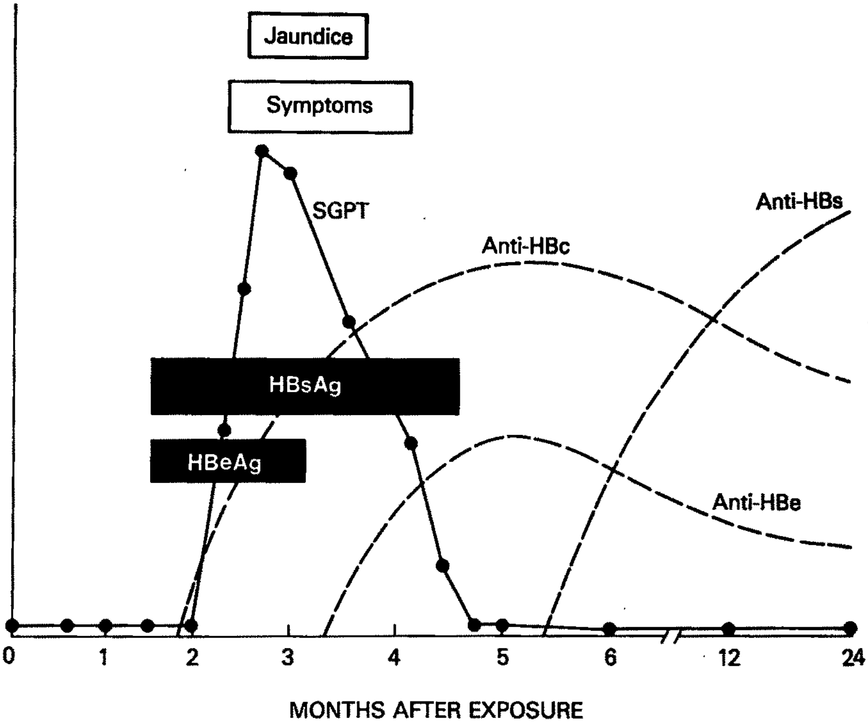


Figure 2 The serologic events associated with the typical course of type B hepatitis. Abbreviations: hepatitis B surface antigen, HBsAg; hepatitis B e antigen, HBeAg; serum glutamic-pyruvic transaminase, SGPT: antibody to HBsAg, anti-HBs; antibody to hepatitis B core antigen, anti-HBc; antibody to HBeAg, anti-HBe.

serologic tool in epidemiological studies and has been the best single marker to document previous or ongoing hepatitis B virus infection.

HBeAg becomes detectable concurrent with or shortly after HBsAg (1, 12). The presence of this serologic marker generally indicates the presence of high levels of virus, DNA polymerase activity, and intact Dane particles. It is usually shorter-lived than HBsAg, disappearing at or around the peak of clinical and biochemical evidence of acute hepatitis. Antibody to HBeAg (anti-HBe) arises soon after HBeAg disappears. This seroconversion from HBeAg to anti-HBe typically occurs at the height of clinical illness and biochemical evidence of hepatitis. This seroconversion does not occur in persons who ultimately develop the chronic HBsAg carrier state. For this reason, the seroconversion from HBeAg to anti-HBe is a favorable prognostic sign–indicating the start of resolution of clinical illness and viral infection.

In adults, approximately 30–40% of individuals infected with the hepatitis B virus will suffer symptomatic acute type B hepatitis (10). Among these persons, 1–3% will develop fulminant hepatitis with a fatal outcome. From a serologic point of view, there is little that distinguishes persons with fulminant type B hepatitis. Most fulminant cases remain HBsAg-positive until death (although titers of HBsAg may fall), and often these persons lose HBeAg reactivity even as they deteriorate clinically.

Subclinical Hepatitis B Virus Infection

At least half of persons infected with the hepatitis B virus do not develop clinically apparent hepatitis, but rather have a transient subclinical infection (also referred to as a primary antibody response or silent seroconversion) (10). The serologic events that occur in this type of infection are depicted in Figure 3. HBsAg appears in low titer and for a brief period only. Indeed, in many persons this period of antigenemia is not detected despite weekly bleedings. This is followed immediately by the appearance of anti-HBs. Typically, such individuals develop the highest and most sustained titers of anti-HBs. Anti-HBc and anti-HBe also appear, but generally in lower titers than anti-HBs.

Mild and transient elevations in transaminases can occur at or around the time of appearance of HBsAg or anti-HBs, but symptoms are absent or

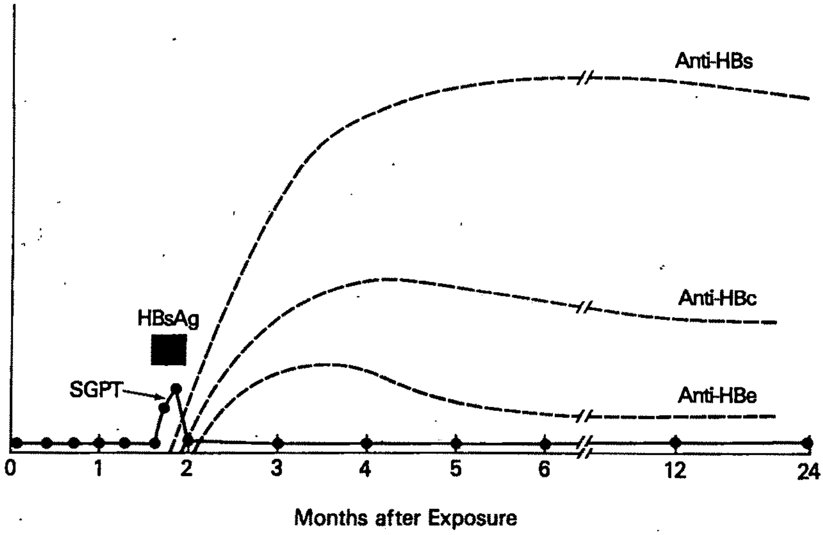


Figure 3 The serologic events associated with subclinical hepatitis B virus infection. Abbreviations: see legend to Figure 2.

extremely mild and nonspecific. This pattern of serologic response is common and it probably accounts for the finding of anti-HBs (with or without anti-HBc) in the sera of persons who deny a history of acute hepatitis.

Development of the Chronic HBsAg Carrier State

Five to ten percent of adults infected with the hepatitis B virus develop the chronic HBsAg carrier state (10). The serologic events that occur with the development of the chronic carrier state are depicted in Figure 4. HBsAg develops during the late incubation period of the disease, rises to high titer (almost always $\geq 100\ \mu g/ml$) and persists for years, or decades, or for life. Anti-HBs does not develop in persons who become HBsAg carriers (although low levels of anti-HBs directed against an opposite serotype are found in a small percentage of carriers). In contrast, anti-HBc appears and rises to very high titer. This anti-HBc consists of both IgG and IgM antibody. Indeed, the finding of high levels of anti-HBc, especially if belonging to the IgM subclass, should suggest the presence of persistent infection.

HBeAg appears with or shortly after HBsAg, rises to high titer and persists (1, 12). Chronic HBsAg carriers typically develop the highest titers of HBeAg (detectable by immunodiffusion as well as radioimmunoassay) (14). However, HBeAg does not persist as long as does HBsAg. The dura-

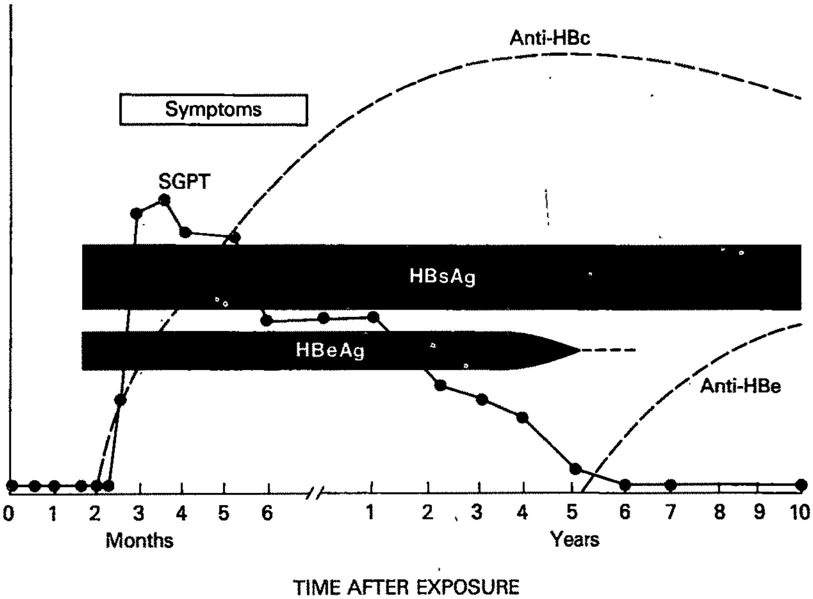


Figure 4 The serologic events associated with the development of the chronic hepatitis B virus carrier state. Abbreviations: see legend to Figure 2.

tion of HBeAg-positivity is variable, lasting for as short as a year to as long as a decade or more. In the majority of persons, however, HBeAg eventually decreases in titer and disappears, to be followed by the development of anti-HBe.

Chronic HBsAg carriers typically have mild or asymptomatic acute disease, which explains why most persons found to be HBsAg-positive deny a history of hepatitis or jaundice. Nevertheless, they do develop elevated transaminases and hepatitis. Furthermore, this hepatitis generally persists and develops into chronic hepatitis. The severity of the chronic hepatitis is variable—from a mild chronic persistent hepatitis that eventually resolves to a severe chronic active hepatitis that rapidly leads to cirrhosis.

The natural history of the HBsAg carrier state has not been well defined. There is a tendency for the chronic liver disease associated with the carrier state to improve with time, resolving spontaneously after many years. Furthermore, the time of resolution of the chronic hepatitis activity often coincides with the seroconversion from HBeAg to anti-HBe. This explains why the HBsAg carrier with HBeAg is likely to have been a carrier for a short period of time and to have some degree of chronic hepatitis. In contrast, the HBsAg carrier with anti-HBe is likely to have been a carrier for many years and to have little or no chronic hepatitis disease activity (a truly "healthy" HBsAg carrier or a person with inactive cirrhosis) (7).

INTERPRETATION OF SEROLOGIC MARKERS OF HEPATITIS B VIRUS INFECTION

Sensitive assays (including radioimmunoassays) are now available for HBsAg, anti-HBs, and anti-HBc. While testing of sequential serum specimens is helpful in diagnosing serologic status, frequently only a single specimen is available and a single set of results requires interpretation (Table 1).

HBsAg may be present without either anti-HBs or anti-HBc (pattern 1). This occurs in only one situation: during the incubation period of acute, type B hepatitis. This finding should suggest that the patient is about to come down with acute, symptomatic type B hepatitis. In all but these very rare cases, HBsAg is present along with anti-HBc (pattern 2). This is the serologic pattern in both acute type B hepatitis and the chronic carrier state. Serologically one cannot distinguish these two possibilities. The diagnosis of the carrier state generally requires the finding of HBsAg persisting for at least six months. The finding of HBsAg should always be considered indicative of ongoing hepatitis B virus infection.

The presence of anti-HBs and anti-HBc together (pattern 3) is the usual

Table 1 Interpretation of serologic markers of hepatitis B virus (HBV) infection

	Serologic reactivity			
Pattern	HBsAg	Anti-HBs	Anti-HBc	Interpretation
1	+	–	–	early (presymptomatic) acute type B hepatitis.
2	+	–	+	(*a*) acute type B hepatitis, or (*b*) chronic HBsAg carrier state
3	–	+	+	recovery from type B hepatitis
4	–	+	–	(*a*) long after HBV infection, or (*b*) immunization with HBsAg
5	–	–	+ ++	(*a*) long after HBV infection, or (*b*) immediate recovery from type B hepatitis (*c*) "low level" carrier state

finding in persons convalescent from type B hepatitis. These persons can be considered fully recovered and immune from further hepatitis B virus infection. Occasionally, persons will be positive for anti-HBs alone, without HBsAg or anti-HBc (pattern 4). Such individuals have probably had hepatitis B virus infection (especially a subclinical infection) long in the past, having since lost detectable anti-HBc reactivity. This pattern of serologic markers is also seen in persons who have been immunized with HBsAg (as with a hepatitis B virus vaccine) rather than infected with the virus.

Finally, individuals may have anti-HBc alone, without either HBsAg or anti-HBs (pattern 5). This pattern can be interpreted in several ways, dependent in part on the anti-HBc titer or antibody subclass—IgG or IgM. First, these individuals may have recovered from type B hepatitis long in the past and have since lost anti-HBs reactivity. In this situation, the anti-HBc titer should be low and should consist entirely of IgG antibody. Second, these persons may have recently recovered from type B hepatitis and be in the "window period" after HBsAg has disappeared and before anti-HBs has become detectable. In this situation, anti-HBc should be present in high titer and should consist partially of IgM antibody. Finally, such persons may be "low level carriers," i.e. chronically infected with the hepatitis B virus but producing HBsAg at too low a level to be detectable. In this situation as well, anti-HBc should be present in high titer and may consist in part of IgM antibody (9). It has been suggested that such "low level carriers" are responsible for cases of type B hepatitis after transfusion with HBsAg-negative blood.

It should also be noted that any of these five patterns may be the result of passive transfer, including blood transfusion, immune serum globulin injection, or in newborns.

HBeAg AND ANTI-HBe TESTING

Testing for HBeAg and anti-HBe is usually helpful only in persons who are HBsAg-positive. Recently, a radioimmunoassay for HBeAg and anti-HBe has become commercially available. This allows for the detection of either HBeAg or anti-HBe in the majority of HBsAg-positive persons (12).

In acute hepatitis, the finding of HBeAg correlates with early and still active disease. Anti-HBe suggests that the infection is beginning to resolve. The seroconversion from HBeAg to anti-HBe usually occurs at the peak of disease activity and guarantees that the patient will not become a chronic HBsAg carrier.

In chronic HBsAg-positive hepatitis, the presence of HBeAg has important clinical correlations (Table 2). Carriers with HBeAg are more likely to be young, to have the recent onset of the carrier state ($<$ 10 years), and to have some degree of chronic hepatitis. In contrast, carriers with anti-HBe are likely to be older, to have had the carrier state for many years, and to have little or no underlying chronic hepatitis activity (7).

Recent interest in HBeAg has centered around the correlation of this antigen with infectivity. All HBsAg carriers should be considered potentially infectious; however, they clearly vary in degree of infectiousness. The presence of HBeAg appears to be a convenient marker for high levels of infectivity, and the presence of anti-HBe for low levels of infectivity. Epidemiological studies first documented the increased infectivity of HBeAg-positive chronic carriers. Thus, chronic HBsAg carrier mothers who are positive for HBeAg regularly transmit type B hepatitis to their newborns, whereas mothers positive for anti-HBe rarely transmit this infection (4, 16). Type B hepatitis is common following an accidental needlestick exposure to blood positive for both HBsAg and HBeAg, but is rare after such exposure to blood containing HBsAg and anti-HBe (8, 19). The association of HBeAg with increased infectivity is also supported by serologic analyses. HBsAg carriers with HBeAg usually have high levels of

Table 2 Clinical and serologic correlates of HBeAg and anti-HBe

Characteristic	HBeAg	Anti-HBe
HBsAg titers	high	low or moderate
DNA-polymerase	present	absent
Dane particles	detectable	rare or undetectable
Duration of carrier state	recent onset	long-standing
Serum transaminases	usually elevated	usually normal
Underlying liver disease	active	inactive or insignificant
Infectivity	high	low

HBsAg, detectable DNA polymerase activity, and intact Dane particles in the serum (15). In contrast, carriers with anti-HBe usually have low or moderate titers of HBsAg, absence of detectable DNA polymerase, and few or no detectable Dane particles. Finally, infectivity titration studies performed in chimpanzees have compared the titers of virus in HBsAg positive sera which contain either HBeAg or anti-HBe. An HBeAg-positive serum was found to be infectious to a dilution of 10^{-8}, whereas an anti-HBe positive specimen (with an equivalent titer of HBsAg) was infectious only to a dilution of 10^{-1} (20). These studies reveal that chronic HBsAg carriers vary considerably in degree of infectivity, and that testing for HBeAg is a relatively easy and seemingly accurate indicator of those carriers with high titers of virus.

CONCLUSION

Sensitive and specific assays are now available for each of the antigens and antibodies related to hepatitis B virus. The finding of HBsAg in the serum, of course, documents the presence of ongoing hepatitis B virus infection. Some estimate of the stage of the infection and of the degree of infectivity of an HBsAg-positive individual can be obtained by also testing for HBeAg and anti-HBe. In the HBsAg-negative individual, previous infection can be documented by the finding of anti-HBc, and some estimate of how recently the infection occurred can be had by testing for anti-HBc titer or antibody subclass. Finally, immunity and convalescence from hepatitis B virus infection can be demonstrated by the finding of anti-HBs. Properly applied, these antigens and antibodies can yield valuable information on the infectivity, stage of disease, and immune status of the patient.

Literature Cited

1. Aikawa, T., Sairenji, H., Furuta, S., Kiyosawa, K., Shikata, T., Imai, M., Miyakawa, Y., Yanase, Y., Mayumi, M. 1978. Seroconversion from hepatitis B e antigen to anti-HBe in acute hepatitis B virus infection. *N. Engl. J. Med.* 298: 439–42
2. Almeida, J. D., Rubenstein, D., Scott, E. J. 1971. New antigen-antibody system in Australia-antigen positive hepatitis. *Lancet* 2:1225–27
3. Barker, L. F., Chisari, F. V., McGrath, P. P., Dalgard, D. W., Kirschstein, R. L., Almeida, J. D., Edgington, T. S., Sharp, D. G., Peterson, M. R., 1973. Transmission of type B viral hepatitis to chimpanzees. *J. Infect. Dis.* 127:648–62
4. Beasley, R. P., Trepo, C., Stevens, C. E., Szmuness, W. 1977. The e antigen and vertical transmission of hepatitis B surface antigen. *Am. J. Epidemiol.* 105: 94–98
5. Blumberg, B. S., Alter, H. J., Visnich, S. 1965. A "new" antigen in leukemia sera. *J. Am. Med. Assoc.* 191:541–545
6. Dane, D. S., Cameron, C. H., Briggs, M. 1970. Virus-like particles in serum of patients with Australia antigen associated hepatitis. *Lancet* 2:695–98
7. Eleftheriou, N., Heathcote, J., Thomas, H. C., Sherlock, S. 1975. Incidence and clinical significance of e antigen and antibody in acute and chronic liver disease. *Lancet* 2:1171–73
8. Grady, G. F., Lee, V. A., Prince, A. M., Gitnick, G. L., Fawaz, K. A., Vyas, G. N., Levitt, M. D., Senior, J. R., Galambos, J. T., Bynum, T. E., Singleton, J.

W., Clowdus, B. F., Akdamar, K., Aach, R. D., Winkelman, E. I., Schiff, G. M., Hersh, T. 1978. Hepatitis B immune globulin for accidental exposures among medical personnel: final report of a multicenter controlled trial. *J. Infect. Dis.* 138:625–38

9. Hoofnagle, J. H. 1980. Specific IgM and IgG antibody responses during type B hepatitis. In *Proceedings of the Symposium: Frontiers of Science and the Liver,* ed. P. D. Berk, T. C. Chalmers. In press
10. Hoofnagle, J. H., Seeff, L. B., Bales, Z. B., Gerety, R. J., Tabor, E. 1978. Serologic responses in hepatitis B. In *Viral Hepatitis,* ed. G. N. Vyas, S. N. Cohen, R. Schmid, 22:219–42. Philadelphia: Franklin Institute. 748 pp.
11. Kaplan, P. M., Greenman, R. L., Gerin, J. L., Purcell, R. H., Robinson, W. 1973. DNA polymerase associated with human hepatitis B antigen. *J. Virol.* 12:995–1005
12. Krugman, S., Overby, L. R., Mushahwar, I. K., Ling, C. M., Froesner, G. G., Deinhardt, F. 1979. Viral hepatitis, type B. Studies on natural history and prevention re-examined. *N. Engl. J. Med.* 300:101–7
13. Magnius, L. O., Espmark, J. A. 1972. New specificities in Australia antigen positive sera distinct from the Le Bouvier determinants. *J. Immunol.* 109:1117–21
14. Nielsen, J. O., Dietrichson, O., Juhl, E. 1974. Incidence and meaning of the "e" determinant among hepatitis-B antigen-positive patients with acute and chronic liver diseases. *Lancet* 2:913–15
15. Nordenfelt, E., Andren-Sandberg, M. 1976. Dane particle–associated DNA polymerase and e antigen: relation to chronic hepatitis among carriers of hepatitis B surface antigen. *J. Infect. Dis.* 134:85–89
16. Okada, K., Kamiyama, I., Inomata, M., Imai, M., Miyakawa, Y., Mayumi, M. 1976. e antigen and anti-e in the serum of asymptomatic carrier mothers as indicators of positive and negative transmission of hepatitis B virus to their infants. *N. Engl. J. Med.* 294:746–49
17. Robinson, W. S. 1980. The hepatitis B virus. Presented at the First US-Japan Symposium on Viral Hepatitis, New York, NY
18. Robinson, W. S., Clayton, D. A., Greenman, R. L. 1974. DNA of a human hepatitis B virus candidate. *J. Virol.* 14:384–88
19. Seeff, L. B., Wright, E. C., Zimmerman, H. J., Alter, H. J., Dietz, A. A., Gelsher, B. F., Finkelstein, J. D., Garcia-Pont, P., Gerin, J. L., Greenlee, H. B., Hamilton, J., Holland, P. V., Kaplan, P. M., Kiernan, T., Koff, R. S., Leevy, C. M., McAuliffe, V. J., Nath, N., Purcell, R. H., Shiff, E. R., Schartz, C. C., Tamburro, C. H., Vlahcevic, Z., Zemel, R., Zimmon, D. S. 1978. Type B hepatitis after needle-stick exposure: prevention with hepatitis B immune globulin. *Ann. Intern. Med.* 88:285–93
20. Shikata, T., Karasawa, T., Abe, K., Uzawa, T., Suzuki, H., Oda, T., Imai, M., Mayumi, M., Mortisugu, Y. 1977. Hepatitis B e antigen and infectivity of hepatitis B virus. *J. Infect. Dis.* 136: 571–76
21. Summers, J., O'Connell, A., Millman, I. 1975. Genome of hepatitis B virus: restriction enzyme cleavage and structure of DNA extracted from Dane particles. *Proc. Natl. Acad. Sci. USA* 72:4597–4601
22. Summers, J., Smolec, J. M., Synder, R. 1978. A virus similar to human hepatitis B virus associated with hepatitis and hepatoma in woodchucks. *Proc. Natl. Acad. Sci. USA* 75:4533–37
23. Yoshizawa, H., Itoh, Y., Simonetti, J. P., Takahashi, T., Machida, A., Miyakawa, Y., Mayumi, M. 1979. Demonstration of hepatitis B e antigen in hepatitis core particles obtained from the nucleus of hepatocytes infected with hepatitis B virus. *J. Gen. Virol.* 42: 513–19

Ann. Rev. Med. 1981. 32:13–28
Copyright © 1981 by Annual Reviews Inc. All rights reserved

IMMUNE COMPLEXES IN HUMAN DISEASES

❖7379

Ralph C. Williams, Jr., M.D.

Department of Internal Medicine, University of New Mexico School of
Medicine, Albuquerque, New Mexico 87131

INTRODUCTION

During the past decade clinical applications of advances in modern immunology have multiplied at an increasing rate. Nowhere has the importance of basic concepts to the understanding of fundamental disease mechanisms been more useful than in the application of knowledge concerning immune complex phenomena to features of internal medicine. Most recently, very sophisticated methods have been developed for estimating quantitative levels of immune complexes in serum or other body fluids. Moreover, many of these methods have been applied to the study of several common or uncommon disease states. The precise significance of elevations of circulating immune complexes in certain disorders is not yet clear; in other diseases, elevations of immune complexes have proved to be of considerable value in clinical management. The present chapter will provide an overview of the general subject with emphasis on methodology and approaches currently available. A number of concrete examples of the possible clinical significance of elevations of immune complexes will be presented to show how this field can apply to practical patient problems.

GENERAL FEATURES OF FORMATION

Any humoral antibody immune response either to foreign or self-antigens must by definition eventually involve the formation of immune complexes. B cells differentiate into immunocytes or plasma cells secreting humoral antibody; this antibody in turn reacts with its respective antigen and in so doing immune complexes are formed. We use the term antibody excess

0066-4219/81/0401-0013$01.00

when the ratio of antibody to antigen in immune complexes is such that antibody molecules are present in excess. As the relative quantities of antigen and antibody approach optimal ratios for maximum precipitation of antigen-antibody complexes, the term equivalence is used. Then in situations of antigen excess, instead of increasing precipitation out of complexes, a relationship usually is established in which soluble, and hence circulating, immune complexes are formed. These quantitative relationships are important in understanding many of the clinical features of immune-complex-mediated injury associated with a number of disease states (1, 2).

One of the most important aspects of immune complex formation is that related to the quality and immunoglobulin class of antibody involved. Thus, IgM antibodies often appear to be related to certain specific types of primary antigens such as Gram negative bacterial lipopolysaccharides or particulate antigens (3, 4). Immune complexes comprised of IgM antibodies are larger and more rapidly cleared than are those formed from IgG or IgA antibodies. One of the most pertinent features of any immune complex is its ability to activate the complement system. Thus immune complexes composed of more than one antigen molecule and cross-linked lattice-wise by several bivalent IgG molecules are often capable of effective complement pathway activation. The same is true for immune complexes composed of IgM antibodies, which in general induce very effective complement component activation. However, immune complexes comprised of antigen and IgA antibody molecules are not effective in complement activation. Thus IgA immune complex deposits within tissues such as are noted with dermatitis herpetiformis (5) or in IgA nephropathy (6, 7) clinically are not associated with intense complement cascade activation or the other amplifying inflammatory effects associated with such sequelae.

Besides actual immunoglobulin class and physical size of immune complexes, one of the most important fundamental features of their physiology within the body relates to methods of their disposal. A number of features are of interest in this regard. The reticuloendothelial system and particularly the hepatic Kupffer cells provide built-in clearing systems for circulating complexes. Also of great importance is the presence of Fc receptors on a number of circulating blood elements such as monocytes, macrophages, lymphocytes (8), polymorphonuclear cells (9), and even platelets (10). Of even greater potential physiologic importance is the presence of actual receptors for activated complement components or Fc portions of immunoglobulins in various tissue sites. Naturally occurring receptors for activated complement (C3b receptors) have now been described in normal human glomeruli (11) as well as in other areas such as the choroid plexus (12, 13). In addition, naturally occurring built-in tissue receptors for the Fc portion of IgG have been described in the interstitium of the kidney (14) as well as in a number of other specific tissues including the synovium and the lamina propria of the gut (15). It is still unclear why nature has endowed certain anatomic areas with naturally occurring receptors for immune complexes. The case of the renal glomerulus presents a real enigma since this structure represents an area particularly vulnerable to immune complex deposition and injury.

Also of importance is the fact that certain tumors appear to be capable of induction of Fc receptors in tissues during the neoplastic process (16, 17). Since a number of disseminated malignancies such as carcinoma of the breast or colon are associated with marked elevations of levels of circulating complexes (18–20), the tumor-associated induction of endogenous Fc receptors in such circumstances may represent an important way by which tumors evade natural host defenses. Several ways in which this could happen are depicted in Figure 1.

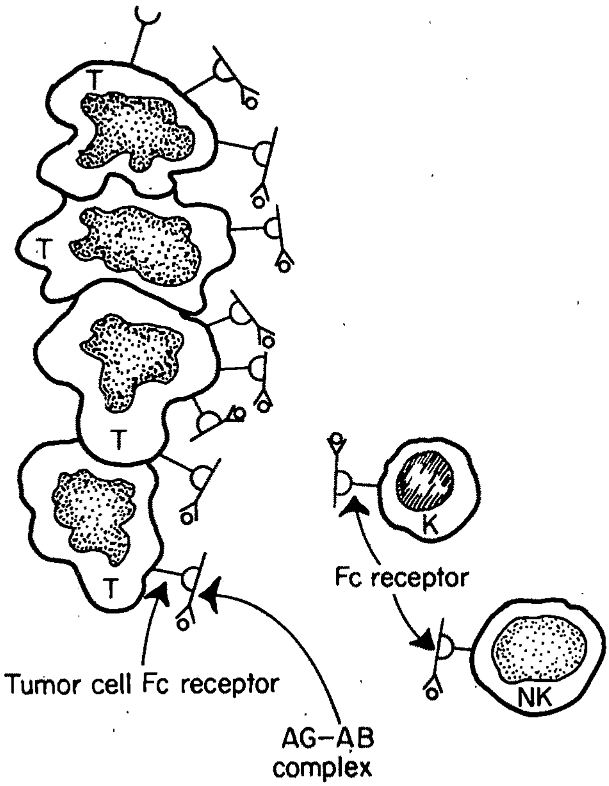


Figure 1 Tumor cells expressing heavy concentration of Fc receptors that have been saturated with antigen-antibody complexes (AG-AB). Natural killer (NK) cells or killer cells (K) are unable to react properly with tumor cells since AG-AB complexes block access to target tumor cells. Alternatively, large quantities of antigen-antibody complexes could also saturate Fc receptors on NK or K cells.

DETECTION METHODS

A number of methods have been developed to measure circulating immune complexes in serum and other body fluids. Historically perhaps the first and most widely applied method for examining the direct effects of immune complex deposition in specific disease states was immunofluorescence (21, 22). Biopsy or autopsy material provides a ready source of tissue that can be directly examined for immune complex constituents such as host immunoglobulin, complement components, and, in some cases, actual putative antigens. Application of immunofluorescence techniques to renal biopsy material was a natural outgrowth of experimental studies in animals directed at immunopathogenesis of experimental serum sickness or Masugi nephritis (21, 23). In many ways the granular lumpy-bumpy immunoglobulin and complement deposits characteristic of immune complex deposits in renal glomeruli during experimental serum sickness or active lupus nephritis have been established as the histologic counterpart of typical lesions mediated by immune complexes. A partial listing of various human disorders in which specific antigens within immune complex reactants have been identified in renal biopsy or other tissue locations is shown in Table 1. Granular immunoglobulin and complement deposition is now accepted as evidence of local tissue immune complex deposits and tissue injury. Surprisingly, however, in only a few human disorders studied to date have putative antigens also been identified by immunologic techniques in similar lesions. These diseases include systemic lupus erythematosus (SLE) (24), hepatitis B–associated vasculitis (25), and, in some patients with infective endocarditis (26, 27), malarial nephritis, or nephritis associated with schistosomiasis (28–30).

In contradistinction to direct immunofluorescence studies of tissue, the methods now available for measuring quantitative amounts of immune complexes in serum have recently multiplied and been refined to an extremely high level of sensitivity (31, 32). These methods fall into several distinct groups; many depend on the reaction of assay systems with activated complement components such as C3b or C1q in the immune complexes themselves. Some of the methods currently available for clinical use are shown in Table 2. Many of the current methods are extremely sensitive: quantitative estimations available using radioimmunoassay are accurate at the 2–5 μg/ml range. Methods that employ C1q, the initially reacting complement component (36, 37), take advantage of the fact that immune complexes (particularly those comprised of IgG or IgM antibodies) activate complement and frequently show activated complement components attached to their surfaces. Use of the conglutinin binding assay (40) also depends on this feature since the radioimmunoassay employing conglutinin,

Table 1 Human disease states in which specific antigens have been directly identified within immune complex deposits

Disease category	Antigen identified in immune complexes or immune complex deposits
Connective tissue disorders	
Systemic lupus erythematosus	DNA, (in glomerular nucleoprotein eluates)
Post-streptococcal glomerulonephritis	Group A strep. (by immunofluorescence in kidney biopsy)
Viral, bacterial, and parasitic diseases	
Hepatitis B–associated polyarteritis or vasculitis	Hepatitis B antigen (identified by EM and immunofluorescence in tissue lesions)
Epstein-Barr (EB) virus-related Burkitt's lymphoma	EB virus antigen (identified in tissue lesions by immunofluorescence)
Infective endocarditis associated with *S. aureus, Streptococcus viridans*	*S. aureus* or *S. viridans* antigens (identified in glomerular biopsy or autopsy material by immunofluorescence)
Nephrotic syndrome associated with secondary syphilis	*Treponema pallidum* antigen (identified by immunofluorescence in kidney biopsy and glomerular eluates)
Typhoid fever	Gram negative Vi antigen (identified by immunofluorescence of renal biopsy material)
P. malariae, associated with nephrotic syndrome	*P. malariae* antigen (identified by immunofluorescence within renal biopsy)
Schistosomiasis, associated with nephrotic syndrome	*S. mansoni* antigen (identified by immunofluorescence in renal biopsy or autopsy material)

a naturally occurring bovine antibody to activated C3, also reacts with complement components attached to immune complexes.

A second major group of immune complex assays utilizes the fine specificity of various cell surface receptors as a probe for detection of immune aggregates. One of the most widely popularized assays involves the use of Raji cells, which are cultured human B cell lines showing an avid receptor for activated C3b or C3d (38, 39). The binding of immune complexes containing C3b or C3d can be utilized in an ultrasensitive radioimmunoassay. Examples of immune complex binding by cultured Raji cells are shown in Figure 2.

The third major type of analysis that has been utilized in examining immune complexes in clinical disease states relates to the physical behavior or size of such complexes. Shifts in size distribution of immunoglobulin, complement or antigen after density gradient or gel filtration separations have also proven useful in identification of immune complex reactants (38). Thus the finding of C3 antigens or IgG in high molecular weight materials separated from serum or fluids suspected of containing complexes has often

Table 2 Commonly used methods for the clinical detection of immune complexes

Method (reference)	Comments
Anticomplementary activity of serum (33)	False positives possible if *in vitro* aggregates of γG occur
Platelet aggregation (34, 35)	Reaction may be inhibited by presence of rheumatoid factors
C1q reactions (36, 37)	Extremely sensitive (3 to 5 μg/ml) when used in radioimmunoassay; non-immune complex materials may react directly with C1q
Raji cell (38, 39)	Sensitive (5–25 μg/ml); may show greater degree of sensitivity for larger (> 11S) complexes
Binding to bovine conglutinin (40)	Also very sensitive (5 μg/ml). Requires C activation and C3d bound to immune complex; may be nonspecifically inhibited by aceto-amido sugars
Rheumatoid factor binding (41)	Can be adapted to radioimmunoassay; does not require prior C activation by complexes
Complexing with protein A of the staphylococcus (42)	Reactions markedly influenced by reactant concentration. No reactivity with IgG3 H chain subgroup
Formation of cryoprecipitate or cryoglobulin (43)	May not correlate with quantitative results in Raji cell or C1q binding; useful in screening

been used in clinical studies to document the presence of immune complex materials.

APPLICATION OF IMMUNE COMPLEX DETERMINATIONS

The number of clinical conditions in which immune complex assays have now been applied is legion. Rather than enumerating all such instances it seems more worthwhile to focus on a selected number of relevant examples to provide useful insight into the current state of the art.

Among infectious disease states infective endocarditis represents a striking example. During the course of chronic bacterial endocarditis, prolonged and potent seeding of the bloodstream with bacterial antigens as well as antigen-antibody complexes occurs. Levels of circulating immune complex during infective endocarditis have been documented by Bayer et al (44) and are often markedly elevated. Actually the measurement of quantitative amounts of such complexes may be clinically useful in distinguishing transient bacteremia from the prolonged sepsis of infective endocarditis (45). A

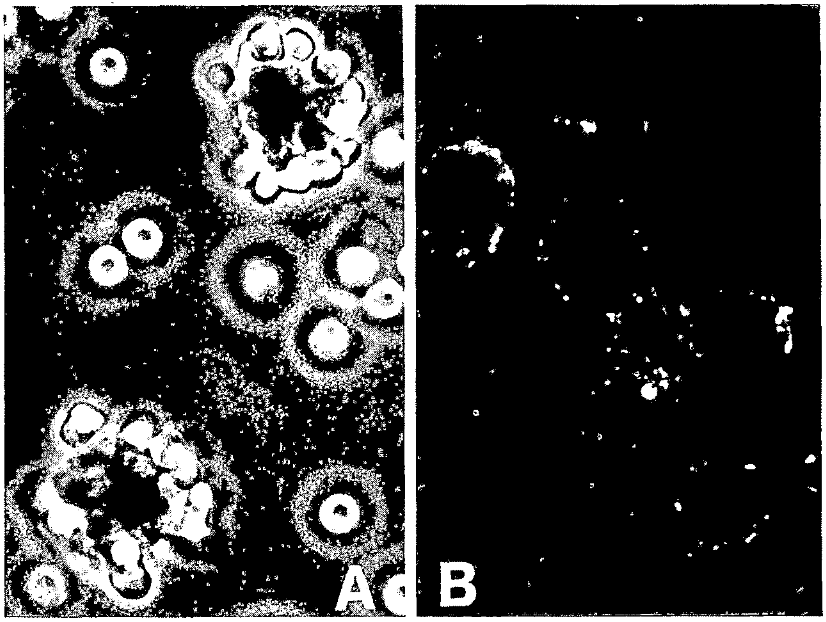


Figure 2 Panel A on left shows cultured Raji cells binding to sensitized erythrocytes coated with IgG antibody and activated complement (C3b). Panel B shows immunofluorescence demonstration of immune complex binding to Raji cells stained with fluorescein-conjugated anti-human IgG. Particulate surface staining for IgG within immune complexes is shown.

number of other common bacterial infections such as disseminated gonococcal infection (46), meningococcal sepsis (47), or pneumococcal pneumonia (48) have now been associated with circulating immune complexes. In the case of meningococcal disease the acute synovitis has been directly linked to a demonstration of immune complexes in synovial inflammatory cells (49).

Parasitic infection as in malaria, trypanosomiasis, or schistosomiasis has repeatedly been linked with the presence of high levels of circulating immune complexes (50–54) as well as the direct tissue-damaging effects of the latter in immune complex renal disease (28–30, 55, 56). The clinical situation in chronic parasitic infection appears to favor unbridled waves of circulating immune complex release since parasitic forms are notoriously facile in changing their external antigenic makeup—thus facilitating cycles of new and different immune complexes with each alteration in surface antigens (57, 58).

Chronic viral infections represent a unique example of the importance of circulating immune complexes in human disease. Perhaps the most widely studied and best documented of the viral diseases is hepatitis B and the various forms of disseminated vasculitis associated with it. Numerous examples of periarteritis-like illness characterized by vascular as well as glomerular immune deposits of hepatitis B antigen and endogenous antibody have now been recorded (25, 59, 60). This clinical picture is rather unique in that well-characterized antigen has also been repeatedly identified within immune deposit lesions. Clinically convincing serial studies of complement profiles as well as peripheral manifestations such as urticaria and synovitis (61, 62) provide further direct evidence for the importance of immune-complex-mediated phenomena during hepatitis B infection.

Perhaps in no other major group of diseases have demonstration and direct study of immune complexes been more useful than in the connective tissue disorders. In rheumatoid synovial inflammation, the inflammatory process is fundamentally driven by local generation and deposition of immune complexes (63–66). Rheumatoid synovial fluids often contain 19S and larger immune aggregates of 7S IgG, and IgM rheumatoid factor as well as activated complement components (64). Such immune aggregates are ingested by local resident phagocytic cells and create more inflammation by subsequent lysosomal degranulation and mediator release. Perhaps the most extensively studied example of immune complex injury within the connective tissue diseases is SLE (1, 24). A typical example of immune complex nephritis associated with active disease is shown in Figures 3*a* and 3*b*. In the case of SLE the documentation has been particularly convincing since eluates of affected glomeruli show not only marked relative quantitative concentration of immunoglobulin with anti-nuclear antibody activity but also the presence of DNA and other strategic nuclear antigenic materials (67).

Recently developed sensitive radioimmunoassays have now been applied to studies of many patients with connective tissue diseases including, of course, rheumatoid arthritis (68), SLE (36, 37), juvenile rheumatoid arthritis (69), mixed connective tissue disease (70), and acute rheumatic fever (71). From studies reported to date it seems clear that different tests used to assay for circulating complexes must show unique specificity for certain types of complexes since quantitative test results using two or three assay systems in parallel do not always show the same degree of positivity or reaction profile. One finding that has emerged from the literature relates to the findings in SLE where a variety of assay systems reveal marked quantitative elevations of immune complexes during active disease.

A fascinating finding has also emerged during experimental work aimed at understanding pathogenetic mechanisms in both SLE and rheumatoid arthritis: Certain elements of dense connective tissue appear to be inordinately susceptible to trapping or adsorption of immune complex reactants. Thus, immune reactants appear to be capable of long-term localization

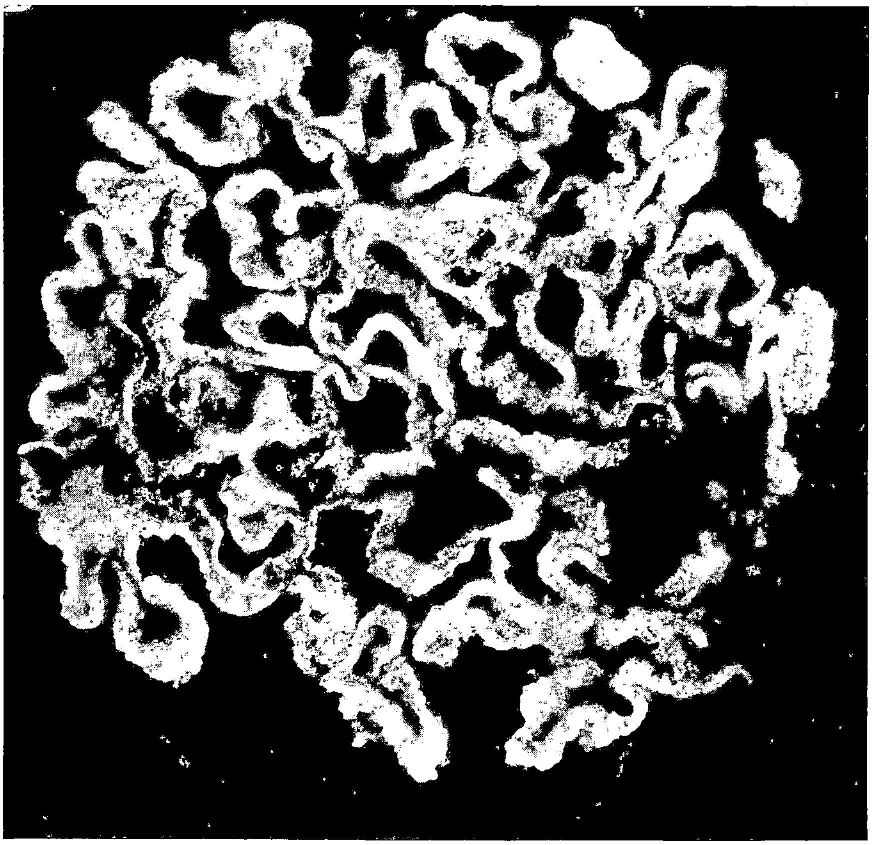


Figure 3a Immunofluorescence demonstration of immune complex deposition within glomeruli affected in systemic lupus erythematosus nephritis. Note the heavy basement membrane staining for IgG.

within dense connective tissue elements (cartilage or tendons) in joints (72, 73). Likewise, experimental protocols by Izui et al (74) have demonstrated adsorption of native DNA to collagen-like materials within glomerular basement membrane. These latter observations suggest that circulating DNA in SLE may first bind to glomerular collagen-like structures and then absorb anti-DNA antibody from the circulation. It seems clear that other similar physical factors may play an important role in tissue site localization of immune complex materials in a number of other diverse disease states.

Immune complexes comprised of tumor-associated antigens and endogenous antibodies generated by the host may be of considerable importance in upsetting the critical balance between tumor and host. Serum-blocking factors amply demonstrated and extensively studied in a number of experi-

G 2-1249

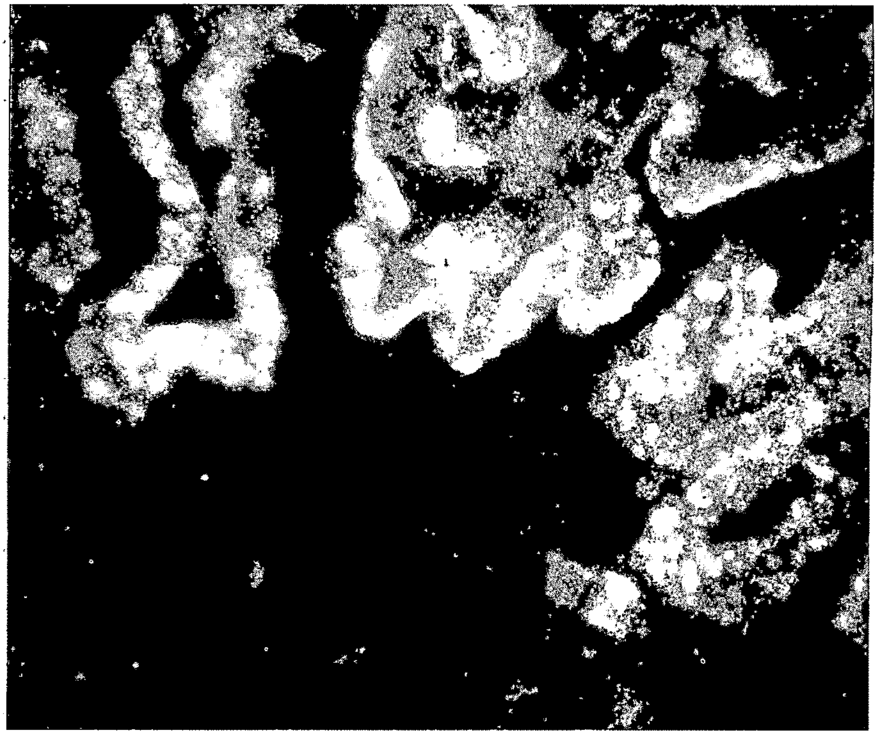


Figure 3b Immune complex deposition as in Figure 3*a*. The granular peripheral deposits of IgG in this patient are shown at magnifications higher than in Figure 3*a*.

mental animal models (75–79) may be of great physiologic importance in allowing the tumor to escape proper immune surveillance. A considerable body of direct evidence using several sensitive assay techniques has now documented marked elevations of circulating immune complexes in various groups of patients with widely differing neoplasms (19, 20, 80–83). A direct correlation, for instance, has been noted between levels of circulating immune complexes at initial diagnosis and ultimate prognosis in patients with leukemia (84). Recently similar studies have also been completed with regard to neuroblastoma (85) and in patients with carcinoma of the colon (20).

An interesting phenomenon in patients with cancer, which has in the past been attributed to immune-complex-mediated renal injury, is the association of subclinical nephritis or overt nephrotic syndrome with various tumors. A role for immune complexes composed of host Ig and complement together with carcino-embryonic antigen was supported by the patient initially reported by Costanza et al (86) where these immune reactants were directly demonstrated in renal biopsy material. Nephrotic syndrome associated with other tumors, particularly malignant lymphomas or Hodgkin's disease (87), may be related to other as yet poorly defined factors since direct documentation of immune complexes (in the form of renal biopsy or autopsy immunofluorescence studies) has not been as uniform. Cell-mediated mechanisms may be of equal importance in the etiology of this latter phenomenon, as suggested by Moorthy et al (88).

If immune-complex-mediated blocking of fundamental host defense mechanisms is important in tumor immunity, it may eventually be clinically worthwhile to take steps to rid tumor-bearing patients of complexes by such methods as plasmapheresis or extracorporeal immunoabsorption. To date there is not a large body of experience or data documenting the feasibility or practical results of such an approach.

A number of miscellaneous human disease states have now been studied with an eye toward relating immune complexes to distinct clinical phenomena. The acute synovitis that sometimes complicates jejuno-ileal anastomosis for morbid obesity has been directly linked to immune complexes composed of host antibody and enteric antigens absorbed through the gut (89, 90). A number of observations have documented marked elevations of immune complex materials during the active phases of inflammatory bowel disease (91, 92). A correlation with some of the extra-gastrointestinal tract manifestations of these disorders such as arthritis, pyoderma gangrenosum, or iritis has been suggested in several studies. A number of patients with chronic active liver disease including Laennec's cirrhosis, chronic active hepatitis, and biliary cirrhosis show pronounced quantitative elevations of circulating immune complexes (93, 94). It is not clear whether such observations are important in assigning immune complexes a direct pathogenetic role or whether they represent secondary phenomena.

Skin disorders of heterogeneous types have also been associated with documented elevations of circulating complexes as well as focal immune deposits, particularly in the basal or junctional areas of the dermis. Examples of these findings are dermatitis herpetiformis (5), necrobiosis lipoidica diabeticorum (95), and erythema multiforme (96).

CONCLUSION

One of the most important problems related to interpretation of the myriads of disease states now being studied for presence of immune complexes lies in the perception of their real physiologic or pathogenetic importance. Thus, with the current extreme sensitivity of assays being used, in any large group of normal sera there are always a few that appear to contain mild elevations of complexes. Also normal human states such as pregnancy have also been associated with modest elevations of immune complexes (97).

At present one of the most important aspects of the central problem probably relates to how the patient or subject handles circulating complexes. If the reticuloendothelial system is not saturated and is normal, probably small amounts of circulating complexes are cleared rapidly and efficiently. However, in a disease such as lepromatous leprosy, schistosomiasis, or even infective endocarditis, the tissue system for handling and disposing of circulating complexes is saturated, and abnormal amounts of such materials are allowed to circulate and induce microvascular injury. Frank and co-workers recently documented tissue defects in handling circulating complexes in patients with systemic lupus erythematosus (98) and biliary cirrhosis (99). Rapid, easily reproduced assays for individual handling of immune complex material may soon be generally available and should greatly aid in clinical assessment of problems in various patient groups.

Literature Cited

1. Cochrane, C. G., Koffler, D. 1973. Immune complex disease in experimental animals and man. *Adv. Immunol.* 16: 185–264
2. Benacerraf, B., Potter, J. L., McCluskey, R. T., Miller, M. 1960. The pathologic effects of intravenously administered soluble antigen-antibody complexes. II. Acute glomerulonephritis in rats. *J. Exp. Med.* 111:195–200
3. Skidmore, B., Chiller, J., Morrison, D., Weigle, W. 1975. Immunologic properties of bacterial lipopolysaccharides (LPS): correlation between the mitogenic, adjuvant, and immunogenic activities. *J. Immunol.* 114:770–75
4. Torrigiani, G., Roitt, I. M. 1965. The enhancement of 19S antibody production by particulate antigen. *J. Exp. Med.* 122:181–93
5. Yaoita, H., Katz, S. I. 1976. Immunoelectron microscopic localization of IgA in skin of patients with dermatitis herpetiformis. *J. Invest. Dermatol.* 67: 502–6
6. McCoy, R. C., Abramowsky, C. R., Tisher, C. C. 1974. IgA nephropathy. *Am. J. Pathol.* 76:123–44
7. Zimmerman, S. W., Burkholder, P. M. 1975. Immunoglobulin A nephropathy. *Arch. Intern. Med.* 135:1217–23
8. Dickler, H. B. 1976. Lymphocyte receptors for immunoglobulin. *Adv. Immunol.* 24:167–214
9. Messner, R. P., Jelinek, J. 1970. Receptors for human G-globulin on human neutrophiles. *J. Clin. Invest.* 49: 2165–71
10. Israels, E. D., Nisli, G., Paraskevas, F., L. G. Israels. 1973. Platelet Fc receptor as a mechanism for Ag-Ab complex-induced platelet injury. *Thromb. Diath. Haemorrh.* 29:434–44
11. Gelfand, M. C., Frank, M. M., Green, I. 1975. A receptor for the third component of complement in the human renal glomerulus. *J. Exp. Med.* 142:1029–34
12. Nyland, H., Matre, R. 1978. Fcγ receptors in human choroid plexus. *Acta Pathol. Microbiol. Scand. Sect. C* 86: 141–43
13. Braathen, L. R., Førre, O. T., Husby, G., Williams, R. C. Jr. 1979. Evidence for Fc IgG receptors and complement factor C3b receptors in human choroid plexus. *Clin. Immunol. Immunopathol.* 14:284–91
14. Gelfand, M. C., Frank, M. M., Green, I., Shinn, J. L. 1979. Binding sites for immune complexes containing IgG in the renal interstitium. *Clin. Immunol. Immunopathol.* 13:19–29
15. Mizoguchi, Y., Tanimoto, K., Yue, C.-L., Horiuchi, Y. 1979. Immune receptors (IgG-Fc and complement receptors) in normal human organs. *Lab. Invest.* 40:703–7
16. Matre, R. 1976. Specificity of Fcγ receptors in human malignant tissue and normal lymphoreticular tissue. *Scand. J. Immunol.* 5:963–68
17. Matre, R., Tønder, O., Thunold, S., Solhaug, J. H. 1976. Properties of Fcγ

receptors in normal and malignant tissues. *Scand. J. Immunol.* 5:361–68
18. Papsidero, L. D., Harvey, S. R., Snyderman, M. C., Nemoto, T., Valenzuela, L., Chu, T. M. 1978. Characterization of immune complexes from the pleural effusion of a breast cancer patient. *Int. J. Cancer* 21:675–82
19. Steele, G., Sonis, S., Stelos, P., Rittgers, R., Zamcheck, N., Finn, D., Maltz, J., Mayer, R., Lockich, J., Wilson, R. E. 1978. Circulating immune complexes in patients following clinically curative resection of colorectal cancer. *Surgery* 83:648–54
20. Riston, S. S., Rossen, R. D., Fryd, D. S., McKhann, C. F. 1979. Circulating immune complexes in colon cancer patient sera. *Cancer* 43:1320–27
21. Dixon, F. J., Feldman, J. D., Vasquez, J. J. 1961. Experimental glomerulonephritis. The pathogenesis of a laboratory model resembling the spectrum of human glomerulonephritis. *J. Exp. Med.* 113:899–920
22. Lerner, R. A., Glassock, R. J., Dixon, F. J. 1967. The role of antiglomerular basement membrane antibody in the pathogenesis of human glomerulonephritis. *J. Exp. Med.* 126:989–1004
23. Unanue, E. R., Dixon, F. J. 1967. Experimental glomerulonephritis: immunological events and pathgenetic mechanisms. *Adv. Immunol.* 6:1–90
24. Koffler, D., Schur, P. H., Kunkel, H. G. 1967. Immunological studies concerning the nephritis of systemic lupus erythematosus. *J. Exp. Med.* 126:607–24
25. Gocke, D. J., Hsu, K., Morgan, C., Bombardieri, S., Lockshin, M., Christian, C. L. 1970. Association between polyarteritis and Australia antigen. *Lancet* 2:1149–53
26. Keslin, M. H., Messner, R. P., Williams, R. C. Jr. 1973. Glomerulonephritis with subacute bacterial endocarditis. Immunofluorescent studies. *Arch. Intern. Med.* 132:578–81
27. Levy, R. L., Hong, R. 1973. The immune nature of subacute bacterial endocarditis (SBE) nephritis. *Am. J. Med.* 54:645–52
28. Ward, P. A., Kibukamusokie, J. W. 1969. Evidence for soluble immune complexes in the pathogenesis of the glomerulonephritis of quartan malaria. *Lancet* 1:283–85
29. Houba, V. 1975. Immunopathology of nephropathies associated with malaria. *Bull. WHO* 52:199–207
30. Hoshino-Shimizu, S., de Brito, T., Kanamura, H. Y., Canto, A. L., Silva, A. O., Campos, A. R., Reena, D. O., DaSilva, L. C. 1976. Human schistosomiasis: *Schistosoma mansoni* antigen detection in renal glomeruli. *Trans. R. Soc. Trop. Med. Hyg.* 70:492–96
31. Barnett, E. V., Knutson, D. W., Abrass, C. K., Chia, D. S., Young, L. S., Liebling, M. R. 1979. Circulating immune complexes: their immunochemistry, detection, and importance. *Ann. Intern. Med.* 91:430–40
32. Williams, R. C. Jr. 1980. Methods of detection. In *Immune Complexes in Clinical and Experimental Medicine.* Cambridge, Mass: Harvard Univ. Press
33. Cream, J. J. 1973. Anticomplementary sera in cutaneous vasculitis. *Br. J. Dermatol.* 89:555–59
34. Penttinen, K., Vaheri, A., Myllylä, G. 1971. Detection and characterization of immune complexes by the platelet aggregation test. I. Complexes formed *in vitro. Clin. Exp. Immunol.* 8:389–97
35. Penttinen, K., Wager, O., Räsänen, J. A., Myllylä, G., Haadawen, E. 1973. Platelet aggregation and cryo-IgM in the study of hepatitis and immune complex states. *Clin. Exp. Immunol.* 15: 409–16
36. Nydegger, U. E., Lambert, P. H., Gerber, H. and Miescher, P. A. 1974. Circulating immune complexes in the serum in systemic lupus erythematosus and in carriers of hepatitis B antigen. Quantitation by binding to radiolabelled C1q. *J. Clin. Invest.* 54:297–309
37. Zubler, R. H., Lange, G., Lambert, P. H., Miescher, P. A. 1976. Detection of immune complexes in unheated sera by a modified ^{125}I-C1q binding test. Effect of heating on the binding of C1q by immune complexes and application of the test to systemic lupus erythematosus. *J. Immunol.* 116:232–35
38. Theofilopoulos, A. N., Wilson, C. B., Dixon, F. J. 1976. The Raji cell radioimmunoassay for detecting immune complexes in human sera. *J. Clin. Invest.* 57:169–82
39. Theofilopoulos, A. N., Wilson, C. B., Bokisch, V. A., Dixon, F. J. 1974. Binding of soluble immune complexes to human lymphoblastoid cells. II. Use of Raji cells to detect circulating immune complexes in animal and human sera. *J. Exp. Med.* 140:1230–44
40. Eisenberg, R. A., Theofilopoulos, A. N., Dixon, F. J. 1977. Use of bovine conglutinin for the assay of immune complexes. *J. Immunol.* 118:1428–34

41. Gabriel, A. Jr., Agnello, V. 1977. Detection of immune complexes. The use of radioimmunoassays with C1q and monoclonal rheumatoid factor. *J. Clin. Invest.* 59:990–1001
42. McDougal, J. S., Redecha, P. B., Inman, R. D., Christian, C. L. 1979. Binding of immunoglobulin G aggregates and immune complexes in human sera to *staphylocci* containing protein A. *J. Clin. Invest.* 63:627–36
43. Levo, Y., Gorevic, P. D., Kassab, H. J., Franklin, D. Z., Franklin, E. C. 1977. Association between hepatitis B virus and essential mixed cryoglobulinemia. *N. Engl. J. Med.* 296:1501–4
44. Bayer, A. S., Theofilopoulos, A. N., Eisenberg, R., Dixon, F. J., Guze, L. B. 1976. Circulating immune complexes in infective endocarditis. *N. Engl. J. Med.* 295:1500–5
45. Bayer, A. S., Theofilopoulos, A. N., Tillman, D. B., Dixon, F. J., Guze, L. B. 1979. Use of circulating immune complex levels in the sero-differentiation of endocarditic and nonendocarditic septicemias. *Am. J. Med.* 66:58–62
46. Walker, L. C., Ahlin, T. D., Tung, K. S. K., Williams, R. C. Jr. 1978. Circulating immune complexes in disseminated gonorrheal infection. *Ann. Intern. Med.* 89:28–33
47. Davis, J. A. S., Peters, N., Mohammed, I., Major, G. A. C., Holborow, E. J. 1976. Circulating immune complexes in a patient with meningococcal disease. *Br. Med. J.* 1:1445–46
48. Tung, K. S. K., Reed, W. P., Williams, R. C. Jr. 1978. Circulating immune complexes during acute pneumococcal infection. Unpublished observations
49. Greenwood, B. M., Whittle, H. C., Bryceson, A. D. M. 1973. Allergic complications of meningococcal disease. II. Immunological investigations. *Br. Med. J.* 2:737–43
50. June, C. H., Contreras, C. E., Perrin, L. H., Lambert, P. H., Miescher, P. A. 1979. Circulating and tissue-bound immune complex formation in murine malaria. *J. Immunol.* 122:2154–61
51. Houba, V., Lambert, P. H., Voller, A., Soyanwo, M. A. O. 1976. Clinical and experimental investigation of immune complexes in malaria. *Clin. Immunol. Immunopathol.* 6:1–12
52. Bout, D., Santoro, F., Carlier, Y., Bina, J. C., Capron, A. 1977. Circulating immune complexes in schistosomiasis. *Immunology* 33:17–22
53. Madwar, M. A., Voller, A. 1975. Circulating soluble antigens and antibody in schistosomiasis. *Br. Med. J.* 1:435–36
54. Ingram, D. G., Soltys, M. A. 1960. Immunity in trypanosomiasis. IV. Immunoconglutinin in animals infected with *Trypansoma brucei. Parasitology* 50:231–39
55. Falcao, H. A., Gould, D. B. 1975. Immune complex nephropathy in schistosomiasis. *Ann. Intern. Med.* 83: 148–54
56. Nagle, R. B., Ward, P. A., Lindsley, H. B., Sadun, E. H., Johnson, A. J., Berkaw, R. E., Hilderbrandt, P. K. 1974. Experimental infections with African trypanosomes. VI. Glomerulonephritis involving the alternate pathway of complement activation. *Am. J. Trop. Med. Hyg.* 23:15–26
57. Wilson, R. A., Barnes, P. E. 1977. The formation and turnover of the membranocalyx on the tegument of *Schistosoma mansoni. Parasitology* 74:61–72
58. Brown, K. N., Brown, I. N. 1965. Immunity to malaria: antigenic variation in chronic infections of *Plasmodium knowlesi. Nature* 208:1286–88
59. Prince, A. M., Trepo, C. 1971. Role of immune complexes involving SH antigen in pathogenesis of chronic active hepatitis and polyarteritis nodosa. *Lancet* 1:1309–12
60. London, W. T. 1977. Hepatitis B virus and antigen-antibody complex diseases. *N. Engl. J. Med.* 296:1528–29
61. Alpert, E., Isselbacher, K. J., Schur, P. H. 1971. The pathogenesis of arthritis associated with viral hepatitis. Complement-component studies. *N. Engl. J. Med.* 285:185–89
62. Onion, D. K., Crumpacker, C. S., Gilliland, B. C. 1971. Arthritis of hepatitis associated with Australia antigen. *Ann. Intern. Med.* 75:29–33
63. Hannestad, K. 1967. Presence of aggregated γ-G-globulin in certain rheumatoid synovial effusions. *Clin. Exp. Immunol.* 2:511–29
64. Winchester, R. J., Agnello, V., Kunkel, H. G. 1970. Gamma globulin complexes in synovial fluids of patients with rheumatoid arthritis. Partial characterization and relationship to lowered complement levels. *Clin. Exp. Immunol.* 6: 689–706
65. Hurd, E. R., LoSpalluto, J., Ziff, M. 1970. Formation of leukocyte inclusions in normal polymorphonuclear cells incubated with synovial fluid. *Arthritis Rheum.* 13:724–33
66. Munthe, E., Natvig, J. B. 1972. Complement-fixing intracellular complexes of IgG rheumatoid factor in rheumatoid

plasma cells. *Scand. J. Immunol.* 1: 217–29

67. Krishnan, C., Kaplan, M. H. 1967. Immunopathologic studies of systemic lupus erythematosus. II. Anti-nuclear reaction of γ-globulin eluted from homogenates and isolated glomeruli of kidneys from patients with lupus nephritis. *J. Clin. Invest.* 46:569–79
68. Halla, J. T., Volanakis, J. E., Schrohenloher, R. E. 1979. Immune complexes in rheumatoid arthritis sera and synovial fluids. A comparison of three methods. *Arthritis Rheum.* 22:440–48
69. Rossen, R. D., Brewer, E. J., Person, D. A., Templeton, J. W., Lidsky, M. D. 1977. Circulating immune complexes and antinuclear antibodies in juvenile rheumatoid arthritis. *Arthritis Rheum.* 20:1485–90
70. Halla, J. T., Volanakis, J. E., Schrohenloher, R. E. 1979. Circulating immune complexes in mixed connective tissue disease. *Arthritis Rheum.* 22:484–89
71. Williams, R. C. Jr., Walker, L. C., Kassaby, M., Mahros, F., Hassaballa, F., Abdin, Z. H., Tung, K. S. K. 1979. Circulating immune complexes in acute rheumatic fever. *J. Clin. Lab. Immunol.* 2:185–90
72. Cooke, T. D., Hurd, E. R., Bienenstock, J., Jasin, H. E., Ziff, M. 1972. The immunofluorescent identification of immunoglobulins and complement in rheumatoid articular collagenous tissue. *Arthritis Rheum.* 15:433–34 (Abstr.)
73. Cooke, T. D., Hurd, E. R., Ziff, M., Josih, H. E. 1972. The pathogenesis of chronic inflammation in experimental antigen-induced arthritis. II. Preferential localization of antigen-antibody complexes to collagenous tissues. *J. Exp. Med.* 135:323–38
74. Izui, S., Lambert, P. H., Miescher, P. A. 1976. *In vitro* demonstration of a particular affinity of glomerular basement membrane and collagen for DNA. A possible basis for a local formation of DNA anti-DNA complexes in systemic lupus erythematosus. *J. Exp. Med.* 144:428–43
75. Hellström, I., Hellström, K. E., Sjögren, H. O. 1970. Serum mediated inhibition of cellular immunity to methylcholanthrene-induced murine sarcomas. *Cell. Immunol.* 1:18–30
76. Takasugi, M., Klein, E. 1971. The role of blocking antibodies in immunological enhancement. *Immunology* 21: 675–84
77. Baldwin, R. W., Price, M. R., Robins, R. E. 1972. Blocking of lymphocyte-mediated cytotoxicity for rat hepatoma cells by tumor-specific antigen-antibody complexes. *Nature (New Biol.)* 238: 185–87
78. Hellström, I., Hellström, K. E., Sjögren, H. O., Warner, G. A. 1971. Serum factors in tumor-free patients canceling the blocking of cell-mediated tumor immunity. *Int. J. Cancer* 8: 185–91
79. Sjögren, H. O., Hellström, I., Bansal, S. C., Hellström, K. E. 1971. Suggestive evidence that the "blocking antibodies" of tumor-bearing individuals may be antigen-antibody complexes. *Proc. Natl. Acad. Sci. USA* 68:1372–75
80. Teshima, H., Wanebo, H., Pinsky, C., Day, N. K. 1977. Circulating immune complexes detected by ^{125}I-C1q deviation test in sera of cancer patients. *J. Clin. Invest.* 59:1134–47
81. Amlot, P. L., Slaney, J. M., Williams, B. D. 1976. Circulating immune complexes and symptoms in Hodgkin's disease. *Lancet* 1:449–51
82. Tsang, K. Y., Singh, I., Blakemore, W. S. 1979. Circulating immune complexes in human osteosarcoma. *J. Natl. Cancer Inst.* 62:743–48
83. Samayoa, E. A., McDuffie, F. C., Nelson, A. M., Go, V. L. W., Luthra, H. S., Brumfield, H. W. 1977. Immunoglobulin complexes in sera of patients with malignancy. *Int. J. Cancer* 19:12–17
84. Carpentier, N. A., Lange, G. T., Fierre, D. M., Lambert, P. H., Miescher, P. A. 1977. Clinical relevance of circulating immune complexes in human leukemia. Association in acute leukemia of the presence of immune complexes with unfavorable prognosis. *J. Clin. Invest.* 60:874–84
85. Brandeis, W. E., Helson, L., Wang, Y., Good, R. A., Day, N. K. 1978. Circulating immune complexes in sera of children with neuroblastoma. Correlation with stage of disease. *J. Clin. Invest.* 62:1201–9
86. Costanza, M. E., Pinn, V., Schartz, R. S., Nathanson, L. 1973. Carcinoembryonic antigen-antibody complexes in a patient with colonic carcinoma and nephrotic syndrome. *N. Engl. J. Med.* 289:520–23
87. Plager, J., Stutzman, L. 1971. Acute nephrotic syndrome as a manifestation of active Hodgkin's disease. Report of four cases and review of the literature. *Am. J. Med.* 50:56–66

88. Moorthy, A. V., Zimmerman, S. W., Burkholder, P. M. 1976. Nephrotic syndrome in Hodgkin's disease. Evidence for pathogenesis alternative to immune complex deposition. *Am. J. Med.* 61:471–77
89. Shagrin, J. W., Frame, B., Duncan, H. 1971. Polyarthritis in obese patients with intestinal bypass. *Ann. Intern. Med.* 75:377–80
90. Wands, J. R., LaMont, J. T., Mann, E., et al. 1976. Arthritis associated with intestinal by-pass procedure for morbid obesity. Complement activation and characterization of circulating cryoproteins. *N. Engl. J. Med.* 294:121–24
91. Jewell, D. P., MacLennan, I. C. M. 1973. Circulating immune complexes in inflammatory bowel disease. *Clin. Exp. Immunol.* 14:219–26
92. Doe, W. F., Booth, C. C., Brown, D. L. 1973. Evidence for complement-binding immune complexes in adult coeliac disease, Crohn's disease, and ulcerative colitis. *Lancet* 1:402–3
93. Thomas, H. C., DeVillers, D., Potter, B., Hodgson, H., Jain, S., Jewell, D. P., Sherlock, S. 1978. Immune complexes in acute and chronic liver disease. *Clin. Exp. Immunol.* 31:150–57
94. Thomas, H. C., Potter, B. J., Sherlock, S. 1977. Is primary biliary cirrhosis an immune complex disease? *Lancet* 2:1261–63
95. Ullman, S., Dahl, M. 1977. Necrobiosis lipoidica. An immunofluorescence study. *Arch. Dermatol.* 113:1671–73
96. Kazmierowski, J. A., Wuepper, K. W. 1977. Erythema multiforme: Immune complex vasculitis of the superficial cutaneous microvasculature. *Clin. Res.* 25:282A
97. Masson, P. L., Delire, M., Cambiaso, C. L. 1977. Circulating immune complexes in normal human pregnancy. *Nature* 266:542–43
98. Frank, M. M., Hamburger, M. I., Lawley, T. J., Kimberly, R. P., Plotz, P. 1979. Detective reticuloendothelial system Fc-receptor function in systemic lupus erythematosus. *N. Engl. J. Med.* 300:518–23
99. Jaffe, C. J., Vierling, J. M., Jones, E. A., Lawley, T. J., Frank, M. M. 1978. Receptor specific clearance by the reticuloendothelial system in chronic liver diseases. *J. Clin. Invest.* 62:1069–77

Ann. Rev. Med. 1981. 32:29–43
Copyright © 1981 by Annual Reviews Inc. All rights reserved

BACTERIAL SURFACE COMPONENTS AND THE PATHOGENESIS OF INFECTIOUS DISEASES

❖7380

Phillip K. Peterson, M.D. and Paul G. Quie, M.D.[1]

Departments of Medicine and Pediatrics, University of Minnesota
Medical School, Minneapolis, Minnesota 55455

Introduction

Host response to bacterial infection is largely governed by characteristics of the bacterial surface (Figure 1). Bacterial surface structures play a major role in the initiation of infection, i.e. in the adherence of bacterial cells to epithelial or endothelial cells of the host, and components of the bacterial surface are involved either directly or indirectly in host injury. Interactions of the bacterial surface with specific cells and humoral elements of the host determine the nature of the host's defense. The inflammatory process itself, a fundamental aspect of host response to bacterial invasion, depends upon the ability of the host to "recognize" bacterial surface components.

In this review article we deal briefly with the above aspects of the interaction of bacterial and host cells. Although our primary interest is in bacterial infections in the human host, it is important to stress that knowledge in this area is incomplete; we must often rely on in vitro observations and on studies in other animal species. Because environmental factors such as nutrient limitations and the presence of antimicrobial agents may profoundly influence the nature of the bacterial surface, results of in vitro experiments must be interpreted with caution. There are also concerns about the integrity and purity of isolated cell surface fractions, the effects of timing and dosage of administered antigens on results, and the variation

---

[1] Dr. Quie is the American Legion Heart Research Professor. This work was supported in part by NIH grant R01 AI 0882-10.

0066-4219/81/0401-0029$01.00

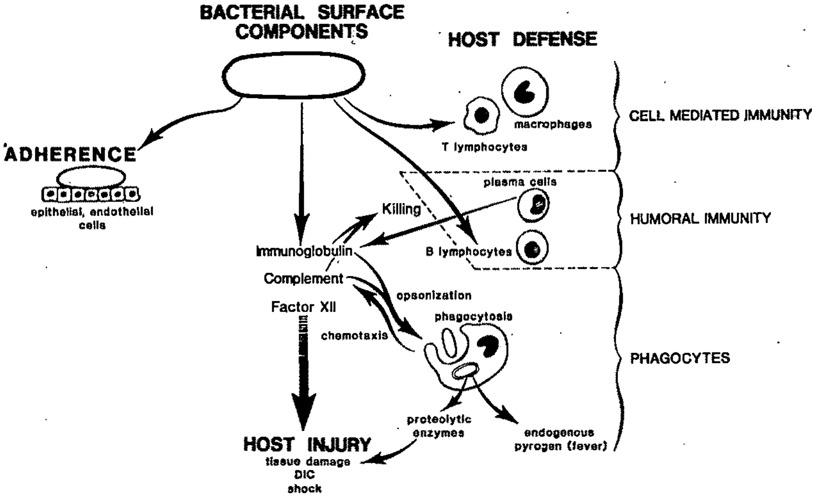


Figure 1 Bacterial surface components and the pathogenesis of infectious diseases.

of responses among animal species. With these important limitations in mind, and given the constraints of the length of this review, we are certain to oversimplify matters.

The Bacterial Cell Envelope

The cell envelopes of gram-positive and gram-negative bacteria are schematically represented in Figure 2. The entire complex set of structures encompassing the cytoplasm of the cell is generally referred to as the bacterial "cell envelope." Although the cell envelope is often regarded as a series of concentric layers, many of its constituents are exposed to the external environment and therefore can be considered part of the cell surface. The differences in gram stainability of microorganisms is a reflection of the different compositions and organizations of their cell walls. This remarkably simple laboratory test can provide us important information regarding the clinical manifestations of infections with these two groups of microorganisms.

The cell envelope not only gives an organism shape, but it is through this structure that a bacterium interacts with its surroundings. The control of nutrient entry into the cell and the exclusion of humoral and antibacterial molecules are governed by the cell envelope (1). Physicochemical characteristics such as surface charge and hydrophobicity significantly affect many of the interactions between the bacterial cell and its environment.

Peptidoglycan, previously referred to as mucopeptide or murein, is present in the cell walls of almost all procaryotes, and it is the only cell wall

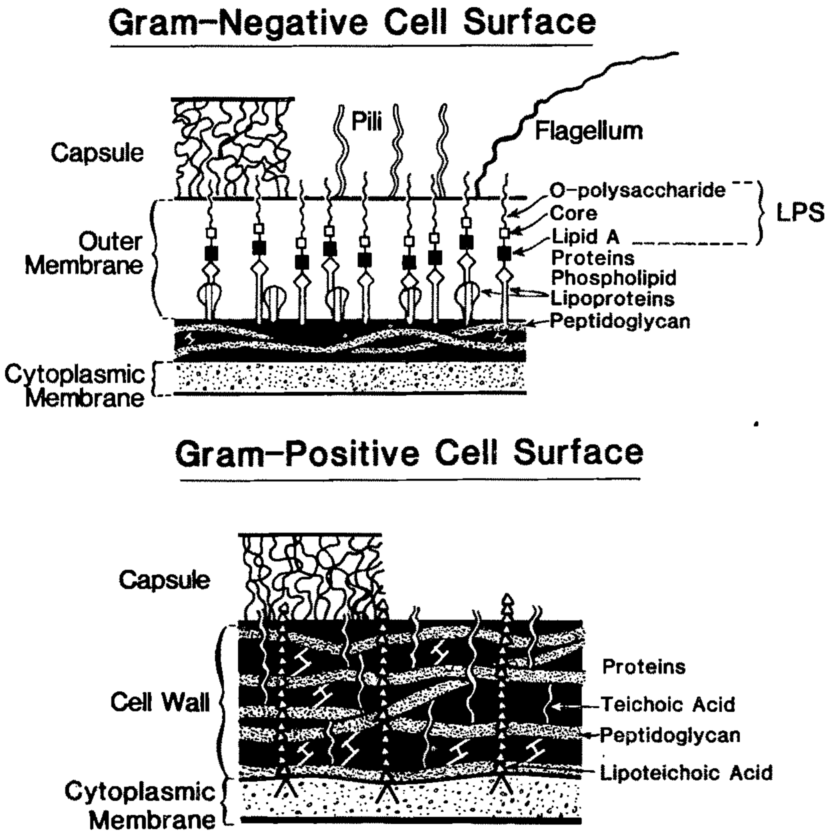


Figure 2 Schematic representation of the cell surfaces of gram-positive and gram-negative bacteria.

component common to both gram-positive and gram-negative bacteria (2). Peptidoglycan comprises about 50% of the cell wall weight of gram-positive bacteria but only about 5–10% of the weight of the gram-negative cell wall. Peptidoglycan is a macromolecular heteropolymer composed of polysaccharide chains cross-linked through short peptides. The polysaccharide moiety (glycan) resembles chitin and is made up of β-1,4 glycosidically linked *N*-acetylglucosamine and *N*-acetylmuramic acid residues.

The teichoic and lipoteichoic acids of the gram-positive cell wall in some ways resemble the lipopolysaccharide (LPS) of the gram-negative cell wall. Lipoteichoic acids are found in most gram-positive bacteria as constituents of the cytoplasmic membrane with extensions of the polar glycerol teichoic acid portion of the molecule through the matrix of the cell wall to the cell surface (3).

The gram-negative bacterial cell wall is relatively complex. The outer membrane is characteristic of gram-negative bacteria, and LPS (endotoxin) is the most unique component of this structure (1). LPS is a heteropolysaccharide chain of repeating oligosaccharide units (the O-polysaccharide or "O" antigen) linked to a central acidic heterooligosaccharide (the core), which is covalently linked to a lipoidal, acetylated glucosamine disaccharide (lipid A). The composition of the O-polysaccharide varies markedly among bacterial species and members of the same species (4); however, the core structure and lipid A are similar in most gram-negative bacteria.

An important component of the bacterial surface of some gram-positive and some gram-negative bacteria is a surrounding capsular polysaccharide. There can be considerable antigenic heterogeneity of this surface structure. For example, there are over 90 capsular (K) antigens found in *Escherichia coli* strains, and more than 80 serotypes of *Streptococcus pneumoniae* with antigenically distinct capsular polysaccharides. Pili (fimbriae) protrude from the cell surface of some gram-negative bacteria, and hair-like structures ("fuzz") are found on the surface of some gram-positive organisms.

Bacterial Adherence

A crucial event in the initiation of many bacterial infections involves the adherence of bacteria to epithelial cells of the host. In some instances, the process of adherence depends on a specific recognition system whereby the infecting microorganism recognizes and discriminates a specific type of host cell receptor (5). Host cell receptor sites interact with specific ligands on the bacterial surface, which results in virtually irreversible adhesion (6).

Smith & Linggood (7) were the first to demonstrate that *E. coli* gastroenteritis in piglets depended on the plasmid-mediated surface antigen K88. This K antigen is actually not a capsular structure but rather is fimbriate or hair-like. Evidence now exists for similar surface structures on many *E. coli* strains that are enterotoxigenic for man (6). Fimbriae appear to initiate colonization of the small intestine, a prerequisite to the development of toxin-mediated diarrhea.

Some enterotoxigenic *E. coli* possess a fimbrial protein (type I pili) that acts as a specific ligand for sugar residues in the epithelial cell of the host (8). This recognition mechanism is blocked by mannose. Adherence of *Vibrio cholerae* appears to occur through an L-fucose-containing receptor (9). There are also mannose-resistant fimbriate colonization factor antigens (CFAI and II) on the surface of many enterotoxigenic *E. coli* strains (6, 10).

Pili also mediate adherence of gram-negative bacteria to epithelial cells of the genitourinary tract (11) (Figure 3). Adherence of *E. coli* to uroepithelial cells is inhibited by mannose and is influenced by a number of

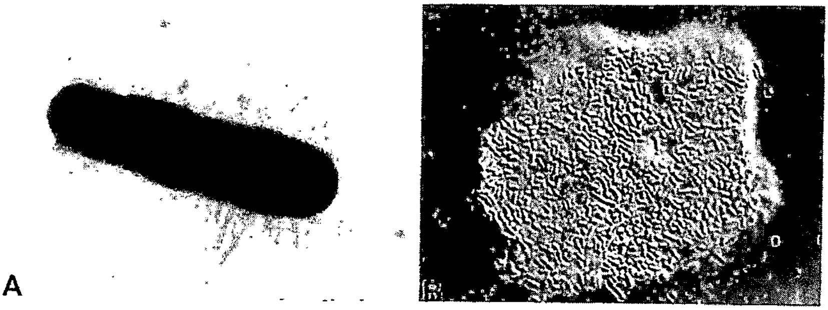


Figure 3 Electronmicrograph of a piliated *E. coli* strain (*A*); bacteria adherent to a urinary tract epithelial cell (*B*). (Courtesy of C. Svanborg Edén)

variables including pH and the menstrual cycle (12). Svanborg Edén & Svennerholm (13) demonstrated that secretory IgA (S-IgA) and IgG antibodies prevent adhesion of *E. coli* to these cells; this suggests a protective role of antibodies to bacterial pili. Several groups of investigators (14, 15) have reported that patients with recurrent urinary tract infections frequently have increased uroepithelial cell receptivity for *E. coli.*

Attachment of *Neisseria gonorrhoeae* (16) and *N. meningitidis* (17) to epithelial cells is likewise dependent upon piliation, and S-IgA and IgG antibodies in human genital secretions provide an inhibitory function (18). Heckels et al (19) suggest that pili of *N. gonorrhoeae* function by overcoming the electrostatic repulsive barrier that exists between the bacterial and host cells.

Another mechanism of bacterial adherence has been described by Costerton et al (20). They showed that in nature bacteria may have a mass of tangled polysaccharide fibers extending from the bacterial surface (Figure 4). This felt-like exopolysaccharide network, or "glycocalyx," may mediate adhesion of bacteria in microcolonies to various surfaces in the natural environment, as well as to certain epithelial surfaces of the human host.

Gram-positive bacteria also have surface components that promote adherence to epithelial cells. Beachey & Ofek (21) demonstrated that adherence of group A streptococci to epithelial cells of the oropharynx is mediated by cell wall lipoteichoic acid. Lipoteichoic acid is present in the hair-like projections on the group A streptococcal surface (Figure 5). Adherence appears to be an important step in the initiation of infection by group A streptococci. Colonization of the vaginal epithelium is important in the pathogenesis of many serious group B streptococcal infections, and recent evidence indicates that the vaginal epithelial cell receptor contains sialic acid residues (22). The bacterial ligand, however, has not been defined.

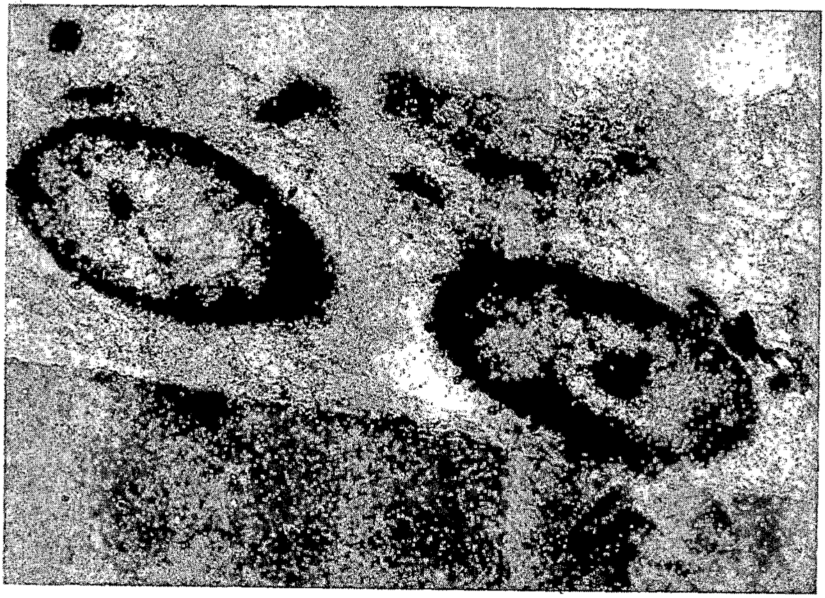


Figure 4 Electronmicrograph showing adherent *Pseudomonas aeruginosa;* a felt-like glycocalyx is seen extending from the bacterial surface. (Courtesy of J. W. Costerton.)

In the case of *Streptococcus pneumoniae,* Selinger & Reed (23) found that these organisms adhere poorly to oropharyngeal cells in vitro and that the bacterial capsule may interfere with adherence.

Adherence is also important in the development of endovascular infections such as endocarditis. Glucan, a dextran that is frequently produced by endocarditis-associated species of streptococci, may contribute to the "stickiness" of these bacteria to endothelial surfaces (24). The initiation of dental caries by certain streptococcal species also appears to depend on glucan but may require cell surface proteins for attachment to the tooth pellicle (25).

New concepts of the prevention and treatment of bacterial infection have emerged from studies on the mechanisms of bacterial adherence (5, 6). The use of substances to block host cell receptor sites, the administration of antibiotics that alter adhesive properties of bacteria, and the development of vaccines using surface fractions such as fimbrial antigens are all under intensive investigation.

Host Injury

Once the epithelial barrier has been breached and access has been gained to the tissues or the systemic circulation of the host, bacteria may preci-

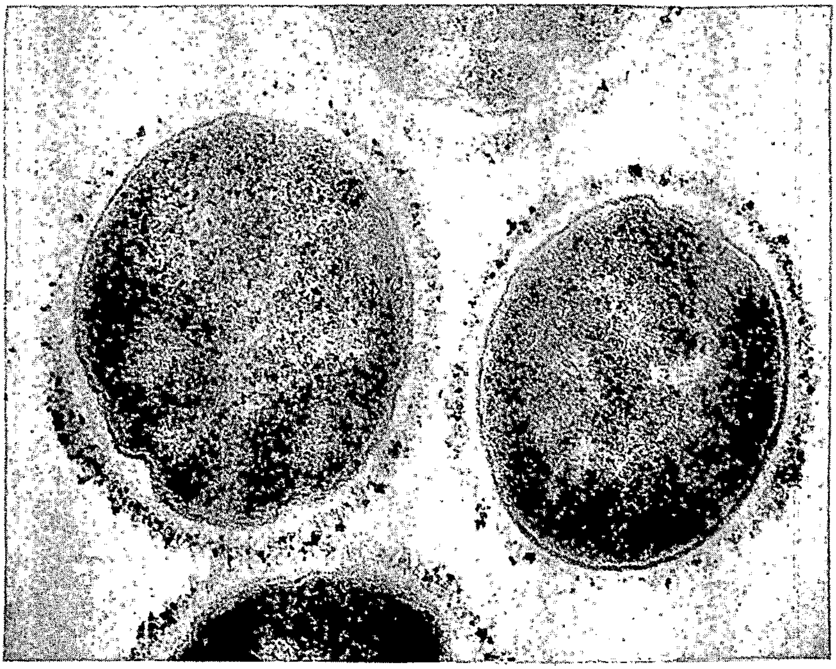


Figure 5 Electronmicrograph of group A streptococci. The hair-like projections ("fuzz") on the surface contain lipoteichoic acid and M protein. (Courtesy of P. P. Cleary.)

pitate significant injury. Although extracellular bacterial products, or exotoxins, are responsible for many of the manifestation of host injury, the bacterial surface can play a major role in this regard. Perhaps the best-understood cell surface component in terms of its potential for causing great damage is the LPS (endotoxin) moiety of the gram-negative outer membrane. Numerous studies have shown that the lipid A component of LPS is responsible for most of its biologic activities, including pyrogenicity, leukopenia, hypotension, and the local Schwartzmann reaction (26).

The lipid A component of LPS is similar in many gram-negative bacteria; however, quantitative or qualitative differences in the composition of LPS may partially explain the variable propensity for the development of the signs of endotoxemia seen with different gram-negative species. For example, the LPS of *Bacteroides fragilis,* an anaerobic gram-negative bacillus, is chemically and biologically distinct from the LPS of aerobic gram-negative bacteria, and *Bacteroides* bacteremia is rarely associated with disseminated intravascular coagulation or purpuric skin lesions (27). Also, a genetic basis for differences in host susceptibility to the endotoxin-properties of LPS has been elegantly demonstrated in the C3H/HeJ mouse (28).

Recognition of the important role of lipid A in the manifestations of endotoxemia has led to the development of vaccines that might protect the host against this highly active molecule. Dramatic cross-protection against a number of gram-negative bacterial species was demonstrated by Braude et al (29) in animals immunized with the J5 mutant of *E. coli* O111. This mutant strain is deficient in an enzyme necessary for incorporation of galactose into LPS to make O-polysaccharide side chains. This approach circumvents the problem posed by the marked antigenic variation among the O antigens of LPS by leaving the common core glycolipid (2-keto-3-deoxyoctonate linked to lipid A) accessible for stimulation of antibodies active against a wide range of gram-negative bacterial core structures. Similar results were reported by McCabe et al (30) using the Re mutant strain of *Salmonella minnesota.* Antibodies against core glycolipid may function primarily by neutralizing the toxic properties of LPS (31).

Components of the gram-positive bacterial cell wall, such as peptidoglycan (32) and lipoteichoic acid (3), also possess endotoxin-like properties; however, these molecules are considerably less potent than LPS. This may explain the relative rarity of shock and purpura seen with gram-positive bacteria. Protein A, a cell wall component of virtually all *Staphylococcus aureus* strains, was recently shown to aggregate platelets (33); this may be important in the pathogenesis of the thromboembolic complications of intravascular staphylococcal infections.

Serious gram-negative and gram-positive bacterial infections are commonly associated with activation of the serum complement system. Many components of the gram-negative (34) and gram-positive (35) cell surface can activate complement, and uncontrolled complement activation may play an important role in host injury. Jacob et al (36) recently demonstrated that C5a production can result in sticking of leukocytes to endothelial surfaces as well as to the aggregation of leukocytes. These events may be critically important in the pathogenesis of leukopenia and in the development of certain clinical states such as "shock lung." Cell wall antigens may also interact with immunoglobulins and complement and result in various forms of injury associated with circulating immune complexes (37).

Host Defense

Successful defense against invading bacteria involves the integrated function of specific humoral and cellular systems of the host, and this aspect of host response is also profoundly influenced by components of the bacterial surface (Figure 1). Immunoglobulins and complement are the most important humoral elements of host defense. Although gram-positive bacteria are resistant to the direct killing action of these serum factors, presumably

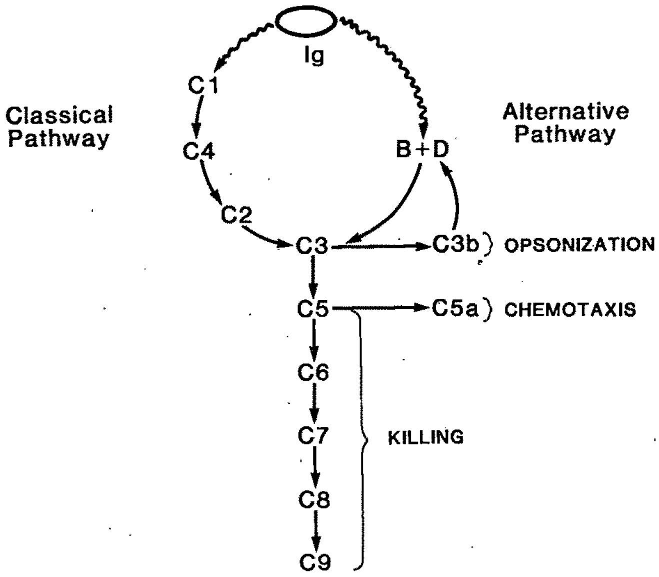


Figure 6 Interactions of the complement system and the bacterial surface.

because of the nature of their peptidoglycan, many gram-negative bacteria are "serum sensitive."

The bactericidal activity of serum depends on activation of the entire complement sequence with formation of the "membrane attack complex" C5b–C9 (Figure 6). Immunoglobulins are important in initiating the complement-mediated killing of many, but not all, gram-negative bacteria (38). There appear to be several determinants of serum sensitivity of gram-negative bacteria including the composition of LPS (1) and the nature of the outer membrane proteins (39). Serum resistance of *E. coli* does not correlate with the presence of capsular antigens (40); however, capsular material of some gram-negative bacteria may impede the killing process (41). Bacterial killing is attended by the disruption of the outer membrane, and, in the presence of serum lysozyme, the peptidoglycan layer is degraded and the bacterial wall is lysed (38).

Serum bactericidal activity seems especially important for host defense against pathogenic *Neisseria* species (42, 43). Gonococcal strains that disseminate from the genital tract are serum-resistant: McCutchan et al (44) proposed that the surface of these organisms possess "receptors" for IgG antibodies that block "natural" bactericidal antibodies found in normal human serum. Also, some patients with severe or recurrent neisserial infections have selective deficiencies of C5, C6, C7, or C8 (45).

Chemotaxis, the directed migration of phagocytic cells along a chemical gradient, is potentially influenced by numerous bacterial surface components capable of activating the complement system with production of the chemotaxin C5a (Figure 6). Chemotaxigenesis by cell surface fractions of *S. aureus*, for example, is mediated by protein A (46), peptidoglycan, teichoic acid, and cell membrane fragments (47).

Phagocytosis of bacteria is generally dependent upon the process of bacterial opsonization. For many organisms, this phenomenon is mediated by the serum complement system with binding of the opsonically active complement fragment C3b to the bacterial surface (Figure 6). C3b and the Fc fragment of immunoglobulin may both function as ligands between the bacterial surface and specific receptors within the plasma membrane of the phagocyte (48). The bacterial surface components that initiate or influence the process of opsonization vary with the bacterial species. The heterogeneity of opsonic requirements found among different species and among strains of the same species most likely reflects differences in their cell surface composition.

Recent studies indicate that peptidoglycan is the key cell wall component in mediating the opsonization of *S. aureus* (49), and the O-polysaccharide of gram-negative bacteria may play a similar role for some gram-negative species (50). Given the immunologic cross-reactivity of bacterial peptidoglycans, it has been postulated that this surface component is an important determinant of "natural immunity" to gram-positive bacteria (51).

The capacity of some bacteria to evade phagocytic cells is an important aspect of microbial defense. Resistance to phagocytosis has long been associated with the presence of cell surface capsular polysaccharides. Encapsulation is an important virulence factor of many species of bacteria that are pathogenic for man. Several groups of investigators have recently suggested that the antiphagocytic characteristic of the bacterial capsule may result from its interference with opsonization by normal (non-immune) serum (52–54), and that this phenomenon may be due to a masking effect by the capsule of wall constituents that would ordinarily mediate opsonization (54, 55) (Figure 7).

Effective opsonization of encapsulated bacteria requires specific anticapsular antibodies, and patients with immunoglobulin deficiency diseases have difficulty primarily with infections due to encapsulated bacteria. Although immunization with polysaccharide vaccines usually results in the development of type-specific immunity, there is also considerable cross-reactivity between some capsular polysaccharides (56, 57).

Cell wall constituents other than the capsule, such as staphylococcal protein A (58) and the M protein of group A *Streptococcus* (59), which is located in the surface "fuzz" (Figure 5), may also interfere with bacterial

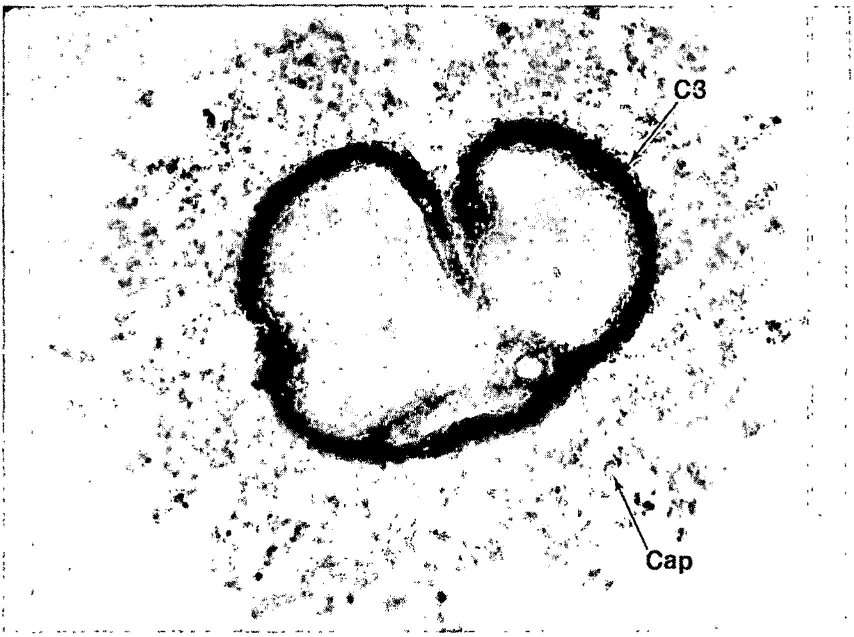


Figure 7 Electronmicrograph of an encapsulated *S. aureus* strain showing the location of C3 on the cell wall beneath the capsular polysaccharide layer (cap). (Courtesy of S. P. Sisson.)

opsonization by normal human serum. Beachey et al (60) extracted highly purified M protein, and the prospects for the development of a streptococcal vaccine using their methods are highly encouraging.

Once phagocytosis has occurred, most bacterial species are rapidly killed within phagocytic vacuoles. Modrzakowski & Spitznagel (61) have recently shown that LPS antagonizes granule cationic proteins. The influence of other bacterial surface components in this and other intracellular microbicidal mechanisms remains a largely unexplored area.

In addition to stimulation of B lymphocytes to produce specific antibodies, a number of bacterial wall constituents may also behave as nonspecific polyclonal B-cell mitogens. Peptidoglycan, protein A, pneumococcal polysaccharide, LPS, and lipoprotein are among these agents (62). The adjuvant activity of some of these cell surface fractions includes stimulation of cell-mediated immune responses. Studies by Chedid et al (63) revealed that the adjuvant activity of Freund's complete adjuvant is due to the minimal adjuvant-active structure of the mycobacterial peptidoglycan *N*-acetyl muramyl-L-alanyl-D-isoglutamine, referred to as MDP (muramyl dipeptide). These investigators found that administration of water-soluble synthetic MDP to animals leads to the development of nonspecific resistance against *Klebsiella pneumoniae* (64). Although agents that promote nonspecific resistance have been shown to stimulate B and T cells and to activate macrophages, the exact mechanism(s) responsible for the activity of nonspecific immunopotentiators is yet to be established.

Conclusions

Research over the past decade has dramatically increased our understanding of the biologic significance of the bacterial surface. The achievements in this realm of biology are founded upon years of basic research into the chemical composition and physical structure of the microbial cell surface. Investigators in this field have deepened our appreciation of many phenomena in the area of bacterial pathogenesis. We now have more understanding of how infection becomes established, how bacteria elicit injury, and how the host and some microbes defend themselves from each other. Knowledge of cell surface structure in relation to biologic function has resulted in innovative approaches to the development of bacterial vaccines and has led to new concepts of the actions of antimicrobial agents. From the perspective of the medical sciences, we have only begun to realize the great promise of this area for the prevention and treatment of human infectious disease.

Literature Cited

1. Costerton, J. W., Ingram, J. M., Chenig, K. J. 1974. Structure and function of the cell envelope of gram-negative bacteria. *Bacteriol. Rev.* 38:87–110
2. Schleifer, K. H. 1975. Chemical structure of the peptidoglycan, its modifiability and relation to the biological activity. *Z. Immunitaetsforsch. Exp. Klin. Immunol.* 149:104–17
3. Wicken, A. J., Knox, K. W. 1977. Biological properties of lipoteichoic acids. In *Microbiology 1977,* ed. D. Schlessinger, pp. 360–65. Washington, DC: Am. Soc. for Microbiol. 593 pp.
4. Lüderitz, O., Staub, A. M., Westphal, O. 1966. Immunochemistry of O and R antigens of *Salmonella* and related *Enterobacteriaceae. Bacteriol. Rev.* 30: 192–255
5. Keusch, G. T. 1979. Specific membrane receptors: Pathogenic and therapeutic implications in infectious diseases. *Rev. Infect. Dis.* 1:517–29
6. Ofek, I., Beachey, E. H. 1980. Bacterial adherence. *Adv. Intern. Med.* 25: 503–32
7. Smith, H. W., Linggood, M. A. 1971. Observations on the pathogenic properties of the K88, HLY and ENT plasmids of *Escherichia coli* with particular reference to porcine diarrhoea. *J. Med. Microbiol.* 4:467–85
8. Ofek, I., Beachey, E. H. 1978. Mannose binding and epithelial cell adherence of *Escherichia coli. Infect. Immun.* 22: 247–254
9. Jones, G. W., Freter, R. 1976. Adhesive properties of *Vibrio cholerae:* Nature of the interaction with isolated rabbit brush border membranes and human erythrocytes. *Infect. Immun.* 14: 240–45
10. Evans, D. G., Evans, D. J. Jr., Tjoa, W. S., DuPont, H. L. 1978. Detection and characterization of colonization factor of enterotoxigenic *Escherichia coli* isolated from adults with diarrhea. *Infect. Immun.* 19:727–36
11. Svanborg Edén, C., Hansson, H. A. 1978. *Escherichia coli* pili as possible mediators of attachment to human urinary tract epithelial cells. *Infect. Immun.* 21:229–37
12. Schaeffer, A. J., Amundsen, S. K., Schmidt, L. N. 1979. Adherence of *Escherichia coli* to human urinary tract epithelial cells. *Infect. Immun.* 24: 753–59
13. Svanborg Edén, C., Svennerholm, A. M. 1978. Secretory immunoglobulin A

and G antibodies prevent adhesion of *Escherichia coli* to human urinary tract epithelial cells. *Infect. Immun.* 22: 790–97
14. Fowler, J. E. Jr., Stamey, T. A. 1977. Studies of introital colonization in women with recurrent urinary infections. VII. The role of bacterial adherence. *J. Urol.* 117:472–76
15. Källenius, G., Winberg, J. 1978. Bacterial adherence to periurethial epithelial cells in girls prone to urinary-tract infections. *Lancet* 1:540–43
16. Swanson, J. 1973. Studies on gonococcus infection. IV. Pili: their role in attachment of gonococci to tissue culture cells. *J. Exp. Med.* 137:571–89
17. McGee, Z. A., Street, C. H., Chappell, C. L., Cousar, E. S., Morris, F., Horn, R. G. 1979. Pili of *Neisseria meningitidis:* effect of media on maintenance of piliation, characteristics of pili, and colonial morphology. *Infect. Immun.* 24:194–201
18. Tramont, E. C. 1977. Inhibition of adherence of *Neisseria gonorrhoeae* by human genital secretions. *J. Clin. Invest.* 59:117–24
19. Heckels, J. E., Blackett, B., Everson, J. S., Ward, M. E. 1976. The influence of surface charge on the attachment of *Neisseria gonorrhoeae* to human cells. *J. Gen. Microbiol.* 96:359–64
20. Costerton, J. W., Geesey, G. G., Cheng, K.-J. 1978. How bacteria stick. *Sci. Am.* 238:86–95
21. Beachey, E. H., Ofek, I. 1976. Epithelial cell binding of group A streptococci by lipoteichoic acid on fimbriae denuded of M protein. *J. Exp. Med.* 143:759–71
22. Zawaneh, S. M., Ayoub, E. M., Baer, H., Cruz, A. C., Spellacy, W. N. 1979. Factors influencing adherence of group B streptococci to human vaginal epithelial cells. *Infect. Immun.* 26:441–47
23. Selinger, D. S., Reed, W. P. 1979. Pneumococcal adherence to human epithelial cells. *Infect. Immun.* 23:545–48
24. Ramirez-Ronda, C. H. 1978. Adherence of glucan-positive and glucan-negative streptococcal strains to normal and damaged heart valves. *J. Clin. Invest.* 62:805–14
25. Staat, R. H., Langley, S. D., Doyle, R. J. 1980. *Streptococcus mutans* adherence: Presumptive evidence for protein-mediated attachment followed by glucan-dependent cellular accumulation. *Infect. Immun.* 27:675–81
26. Galanos, C., Freudenberg, M., Hase, S., Jay, F., Ruschmann, E. 1977. Biological activities and immunogical properties of lipid A. See Ref. 3, pp. 269–76
27. Kasper, D. L., Onderdonk, A. B., Polk, B. F., Bartlett, J. G. 1979. Surface antigens as virulence factors in infection with *Bacteroides fragilis. Rev. Infect. Dis.* 1:278–88
28. Watson, J., Kelly, K., Largen, M. 1977. Genetic and cellular aspects of host response to endotoxin. See Ref. 3, pp. 298–303
29. Braude, A. I., Ziegler, E. J., Douglas, H., McCutchan, J. A. 1977. Antibody to cell wall glycolipid of gram-negative bacteria: induction of immunity to bacteremia and endotoxemia. *J. Infect. Dis.* 136 (Suppl.):S167–73
30. McCabe, W. R., Johns, M. A., Craven, D. E., Bruins, S. C. 1977. Clinical implications of enterobacterial antigens. See Ref. 3, pp. 293–97
31. Young, L. S., Stevens, P., Ingram, J. 1975. Functional role of antibody against "core" glycolipid of *Enterobacteriaceae. J. Clin. Invest.* 56:850–61
32. Heymer, B., Rietschel, E. T. 1977. Biological properties of peptidoglycans. See Ref. 3, pp. 344–49
33. Hawiger, J., Steckley, S., Hammond, D., Cheng, C., Timmons, S., Glick, A. D., Des Prez, R. M. 1979. Staphylococci-induced human platelet injury mediated by protein A and immunoglobulin G Fc fragment receptors. *J. Clin. Invest.* 64:931–37
34. Morrison, D. C., Kline, L. F. 1977. Activation of the classical and properdin pathways of complement by bacterial lipopolysaccharide (LPS). *J. Immunol.* 118:362–68
35. Wilkinson, B. J., Kim, Y., Peterson, P. K., Quie, P. G., Michael, A. F. 1978. Activation of complement by cell surface components of *Staphylococcus aureus. Infect. Immun.* 20:388–92
36. Jacob, H. S., Craddock, P. R., Hammerschmidt, D. E., Moldow, C. F. 1980. Complement-induced granulocyte aggregation: An unsuspected mechanism of disease. *N. Engl. J. Med.* 302:789–94
37. Barnett, E. V., Knutson, D. W., Abrass, C. K., Chia, D. S., Young, L. S., Liebling, M. R. 1979. Circulating immune complexes: their immunochemistry, detection, and importance. *Ann. Intern. Med.* 91:430–40
38. Schreiber, R. D., Morrison, D. C., Podack, E. R., Müller-Eberhard, H. J. 1979. Bactericidal activity of the alternative complement pathway generated

from 11 isolated plasma proteins. *J. Exp. Med.* 149:870–82

39. Hildebrandt, J. F., Mayer, L. W., Wang, S. P., Buchanan, T. M. 1978. *Neisseria gonorrhoeae* acquire a new principal outer-membrane protein when transformed to resistance to serum bactericidal activity. *Infect. Immun.* 20: 267–73
40. McCabe, W. R., Kaijser, B., Olling, S., Uwaydah, M., Hanson, L. Å. 1978. *Escherichia coli* in bacteremia: K and O antigens and serum sensitivity of strains from adults and neonates. *J. Infect. Dis.* 138:33–41
41. Ward, M. E., Lambden, P. R., Heckels, J. E., Watt, P. J. 1978. The surface properties of *Neisseria gonorrhoeae:* determinants of susceptibility to antibody complement killing. *J. Gen. Microbiol.* 108:205–12
42. Goldschneider, I., Gotschlich, E. C., Artenstein, M. S. 1969. Human immunity to the meningococcus. II. Development of natural immunity. *J. Exp. Med.* 129:1327–48
43. Schoolnik, G. K., Buchanan, T. M., Holmes, K. K. 1976. Gonococci causing disseminated gonococcal infection are resistant to the bactericidal action of normal human sera. *J. Clin. Invest.* 58: 1163–73
44. McCutchan, J. A., Katzenstein, D., Norquist, D., Chikami, G., Wunderlich, A., Braude, A. I. 1978. Role of blocking antibody in disseminated gonococcal infection. *J. Immunol.* 121: 1884–88
45. Petersen, B. H., Lee, T. J., Snyderman, R., Brooks, G. F. 1979. *Neisseria meningitidis* and *Neisseria gonorrhoeae* bacteremia associated with C6, C7, or C8 deficiency. *Ann. Intern. Med.* 90: 917–20
46. Harvey, R. L., Kronvall, G., Troup, G. M., Anderson, R. F., Williams, R. C. Jr. 1970. Chemotaxis of polymorphonuclear leukocytes by protein A of the staphylococcus. *Proc. Soc. Exp. Biol. Med.* 135:453–56
47. Schmeling, D. J., Peterson, P. K., Hammerschmidt, D. E., Kim, Y., Verhoef, J., Wilkinson, B. J., Quie, P. G. 1979. Chemotaxigenesis by cell surface components of *Staphylococcus aureus. Infect. Immun.* 26:57–63
48. Verhoef, J., Peterson, P. K., Quie, P. G. 1977. Human polymorphonuclear leucocyte receptors for staphylococcal opsonins. *Immunology* 33:231–39
49. Peterson, P. K., Wilkinson, B. J., Kim, Y., Schmeling, D., Douglas, S. D., Quie, P. G., Verhoef, J. 1978. The key role of peptidoglycan in the opsonization of *Staphylococcus aureus. J. Clin. Invest.* 61:597–609
50. Young, L. S., Martin, W. J., Meyer, R. D., Weinstein, R. J., Anderson, E. T. 1977. Gram-negative rod bacteremia: microbiologic, immunologic, and therapeutic considerations. *Ann. Intern. Med.* 86:456–71
51. Karakawa, W. W., Braun, D. G., Lackland, H., Krause, R. M. 1968. Immunochemical studies on the cross-reactivity between streptococcal and staphylococcal mucopeptide. *J. Exp. Med.* 128: 325–40
52. Peterson, P. K., Wilkinson, B. J., Kim, Y., Schmeling, D., Quie, P. G. 1978. Influence of encapsulation on staphylococcal opsonization and phagocytosis by human polymorphonuclear leukocytes. *Infect. Immun.* 19: 943–49
53. Van Dijk, W. C., Verbrugh, H. A., van der Tol, M. E., Peters, R., Verhoef, J. 1979. Role of *Escherichia coli* K capsular antigens during complement activation, C3 fixation, and opsonization. *Infect. Immun.* 25:603–9
54. Horowitz, M. A., Silverstein, S. C. 1980. Influence of the *Escherichia coli* capsule on complement fixation and on phagocytosis and killing by human phagocytes. *J. Clin. Invest.* 65:82–94
55. Wilkinson, B. J., Sisson, S. P., Kim, Y., Peterson, P. K. 1979. Localization of the third component of complement on the cell wall of encapsulated *Staphylococcus aureus* M: implications for the mechanism of resistance to phagocytosis. *Infect. Immun.* 26: 1159–63
56. Young, L. S., Steven, P. 1977. Cross-protective immunity to gram-negative bacilli: Studies with core glycolipid of *Salmonella minnesota* and antigens of *Streptococcus pneumoniae. J. Infect. Dis.* 136 (Suppl.):S174–80
57. Crumrine, M. H., Fischer, G. W., Balk, M. W. 1979. Immunochemical cross-reactions between type III group B streptococcus and type 14 *Streptococcus pneumoniae. Infect. Immun.* 25:960–63
58. Peterson, P. K.; Verhoef, J., Sabath, L. D., Quie, P. G. 1977. Effect of protein A on staphylococcal opsonization. *Infect. Immun.* 15:760–64
59. Peterson, P. K., Schmeling, D., Cleary, P. P., Wilkinson, B. J., Kim, Y., Quie, P. G. 1979. Inhibition of alternative complement pathway opsonization by

group A streptococcal M protein. *J. Infect. Dis.* 139:575–85

60. Beachey, E. H., Stollerman, G. H., Johnson, R. H., Ofek, I., Bisno, A. L. 1979. Human immune response to immunization with a structurally defined polypeptide fragment of streptococcal M protein. *J. Exp. Med.* 150:862–77
61. Modrzakowski, M. C., Spitznagel, J. K. 1979. Bactericidal activity of fractionated granule contents from human polymorphonuclear leukocytes: Antagonism of granule cationic proteins by lipopolysaccharide. *Infect. Immun.* 25: 597–602
62. Schwab, J. H. 1977. Modulation of the immune response by bacteria. See Ref. 3, pp. 366–73
63. Chedid, L., Audibert, F., Johnson, A. G. 1978. Biological activities of muramyl dipeptide, a synthetic glycopeptide analogous to bacterial immunoregulating agents. *Prog. Allergy* 25:63–105
64. Chedid, L., Parant, M., Parant, F., Lefrancier, P., Choay, J., Lederer, E. 1977. Enhancement of non-specific immunity to *Klebsiella pneumoniae* infection by a synthetic immunoadjuvant (N-acetyl muramyl-L-alanyl-D-isoglutamine) and several analogs. *Proc. Natl. Acad. Sci. USA* 74:2089–93

Ann. Rev. Med. 1981. 32:45–61
Copyright © 1981 by Annual Reviews Inc. All rights reserved

CHLAMYDIAL INFECTIONS[1]

❖7381

J. Schachter, Ph.D.

George Williams Hooper Foundation, University of California,
San Francisco, California 94143

M. Grossman, M.D.

Department of Pediatrics, San Francisco General Hospital,
San Francisco, California 94143

INTRODUCTION

Chlamydia trachomatis has long been recognized as the causative agent of trachoma, a chronic conjunctivitis affecting hundreds of millions of people in developing countries and resulting in millions of cases of blindness (1). Recent studies, however, have focused on the role of *C. trachomatis* in diseases of individuals living in industrialized countries (2, 3). In this milieu most *C. trachomatis* infections are sexually transmitted and they, as other sexually transmitted pathogens, have reached epidemic levels. An estimated 3 million chlamydial infections occur each year in the US, as compared to 2 million gonococcal infections (4). Chlamydial disease occurs in both men and women and there are important consequences to infants with perinatal exposure to chlamydiae (2).

This review focuses on the spectrum of disease associated with *C. trachomatis* (non-lymphogranuloma venereum, non-LGV) infections in industrialized societies. Because the diseases of the genital tract are reviewed extensively elsewhere, a major focus here is the newly developed information relevant to chlamydial infections in infants (2).

---

[1] This work was supported in part by Public Health Service grant EY-02216 from the National Institutes of Health.

0066-4219/81/0401-0045$1.00

THE ORGANISM

The chlamydiae are a group of obligate intracellular bacteria separated into their own order, Chlamydiales (with one family, Chlamydiaceae), on the basis of their unique growth cycle which distinguishes them from all other microorganisms (5). This cycle involves infection of a susceptible host cell by a *Chlamydia*-specified phagocytic process so that these organisms are preferentially ingested (6). Preliminary studies suggest that there may be specific receptor sites on the host cell for different chlamydial agents. For example, *N*-acetyl glucosamine and sialic acid have both been implicated as possible attachment sites for different chlamydiae (7, 8). After attachment and ingestion the chlamydiae remain in a phagosome throughout the growth cycle, but surface antigens of chlamydiae appear to inhibit phagolysosomal fusion (9). These two virulence factors, enhanced ingestion and inhibition of phagolysosomal fusion, speak for an exquisitely adapted parasitism. Once in the cell the chlamydial elementary body, which is the infectious particle, changes to a metabolically active replicating form called the reticulate body (initial body), which synthesizes its own macromolecules and divides by binary fission. The chlamydiae are energy parasites and do not synthesize their own ATP, thus energy-rich compounds must be supplied to them by the host cell (10). As the end of the growth cycle is reached (approximately 48 hours) the initial bodies reorganize into elementary bodies and these particles are released to initiate new infectious cycles.

There are two species within the genus (5). *Chlamydia psittaci* is the causative agent of psittacosis and a common pathogen in avian species and lower mammals. *Chlamydia trachomatis* seems to be a specifically human pathogen (except for a few strains of rodent origin). The two organisms are differentiated on the basis of glycogen-containing inclusions that stain with iodine (*C. trachomatis* is positive while *C. psittaci* is negative in this test) and sensitivity to sulfonamides (*C. trachomatis* is sensitive while *C. psittaci* is resistant).

All chlamydiae share a common genus-specific antigen. The *C. trachomatis* strains also share species- and type-specific antigens to a varying degree. There are 15 serotypes currently recognized (3). Three of these serotypes (L-1, L-2, and L-3) represent the agents causing lymphogranuloma venereum (LGV). LGV strains appear to have different receptor sites and a much broader tissue spectrum in vivo and host spectrum in vitro than the other *C. trachomatis* strains. They should not be included in the same species. The other serotypes of *C. trachomatis* represent the agents causing endemic blinding trachoma (A, B, Ba, and C) and the sexually transmitted *C. trachomatis* strains (D through K), which cause inclusion conjunctivitis, urethritis, cervicitis, etc, and may be capable of causing sporadic eye disease that is clinically indistinguishable from trachoma.

METHODS OF DIAGNOSING *C. TRACHOMATIS* INFECTIONS

The principles of diagnosing chlamydial infection are essentially the same as those of diagnosing any other microbial infection. The agent may be demonstrated by cytology in clinical specimens; serologic means may be used to demonstrate rising antibody titers to chlamydial antigens; or the agent may be isolated from the patient's tissues (11).

Cytology

The unique growth cycle of *Chlamydia* and their intracellular milieu provide specific markers to be used in cytodiagnosis. The demonstration of inclusions (intracellular colonies of chlamydiae) in conjunctival scrapings or genital tract specimens was the method used to elucidate the epidemiology and clinical spectrum of chlamydial infections prior to the 1960s, but these procedures cannot be recommended because of poor sensitivity. Cytology (usually using the Giemsa stain) is still useful in areas where more sophisticated laboratory technology is not available or in some instances where rapid diagnosis can be achieved, but it must be recognized that a failure to demonstrate inclusions will not rule out a chlamydial infection.

Serology

Serology is not particularly useful in diagnosing *C. trachomatis* infections, particularly those of the genital tract (12). There is an abundant immune response to chlamydial infections (3). Virtually all individuals who are not treated very early in the course of their infection will develop a measurable antibody response; cell-mediated immune response is also stimulated (13). Serodiagnosis is made difficult by the very high prevalence of anti-chlamydial antibody in sexually active populations. In addition, many chlamydial infections are chronic, and thus the patient may not be seen at an appropriate time to allow acute and convalescent sera to be collected and show increasing titers. It is more common to find static responses.

The two serologic tests that have been used for chlamydial infection are the complement fixation (CF) test and the microimmunofluorescent (micro-IF) test (3, 14). The CF test is not particularly useful in diagnosing *C. trachomatis* infections other than LGV. The micro-IF test is a much more sensitive test and allows determination of the immunoglobulin class of antibody.

There are not yet enough data to depict the typical antibody response to the systemic infections with *C. trachomatis* (except in infants). For the superficial infections IgM antibodies seem to appear 1–2 weeks postinfection and usually persist for approximately 30 days in the untreated individual. IgG antibodies, which are usually found as early as the IgM antibodies, will persist for years and possibly for life, although the levels may fall with time. Antibodies found in tears or genital tract secretions may be produced locally but more often reflect circulating antibodies (15, 16).

With reinfection IgM antibodies are usually detected and there may or may not be a four-fold increase in IgG titer. Heterotypic second infections occasionally result in rising IgG titers to the original infecting serotype, whereas the IgM response is more specific for the second infecting serotype.

Although the immune response to the chlamydial infection is easily detected, it is not known if the antibodies measured by micro-IF play any role in resistance to reinfection. Some evidence suggests that a relative resistance to reinfection may follow repeated infection. Preliminary studies with trachoma vaccines demonstrated some short-lived protection against infection (3, 17). Unfortunately, a more severe disease occurred in some of the immunized individuals. This observation, together with experimental studies on infecting primates, led to the suggestion that chlamydial diseases are immune-mediated. In our opinion, this conjecture is unwarranted. Primary chlamydial infections obviously can cause disease. Antigen-antibody or cell-mediated immune reactions at the involved sites can add to the disease process, but are not necessarily responsible for the entire process. Treatment with steroids reactivates some inapparent chlamydial infections, causing exacerbations (18, 19), and immunosuppression may predispose a person to systemic chlamydial infection (18, 20).

In selected instances, serology may be useful in diagnosing the infection; for example, in an adult's first acquisition of chlamydiae (paired sera will show rising titers) (21). In some diseases involving systemic complications of superficial chlamydial infections, the exceptionally high titers seen may possibly provide some corroboration for clinical diagnosis. For example, women who develop acute salpingitis from a chlamydial infection will have much higher antibody levels than women with uncomplicated cervicitis (22); men with chlamydial epididymitis will have much higher antibody titers than men with chlamydial urethritis. The rule of thumb is that systemic complications result in higher antibody titers quite similar to the high titers observed in individuals with proven LGV infections. This is seen also in infants with perinatally acquired infection: an infant with pneumonia will have much higher antibody titers than a child developing uncomplicated conjunctivitis (23, 24).

Cultural Methods

The ideal method of diagnosing chlamydial infections is isolation of the organism from the involved site. Because chlamydiae are obligatory intracellular parasites, isolation attempts cannot be performed on artificial media but require a tissue culture facility. The methods of choice involve treating a susceptible cell monolayer with an antimetabolite that interferes with the replication or metabolism of the host cell while allowing chlamydial functions. The two most commonly used methods involve treatment of McCoy cells with either 5-iodo-2-deoxyuridine or cycloheximide (25, 26). The most important single step in the isolation procedure involves enhancing the contact (by centrifiguation) between the host cell and the *Chlamydia* (27). After an appropriate incubation period, which varies from 24 to 72 hours depending upon staining procedures used or the necessity for an early diagnosis, the monolayers are stained and examined microscopically for the presence of inclusions.

Isolation procedures are usually the methods of choice in diagnosing chlamydial infections, but they are not without pitfalls. It is imperative that the involved anatomic site be appropriately sampled. Culture of discharges is inadequate and an adequate sample of involved epithelial cells must be obtained. The sensitivity of chlamydial culture techniques is unknown. Thus, there is no "gold standard" with which one can compare chlamydial serologic procedures. The sensitivity of culture techniques for nongonococcal urethritis or cervicitis is probably about 70–80%. Culture for other conditions (except neonatal conjunctivitis) is likely to be less sensitive. Systemic infections are probably harder to diagnose by isolation procedures because of sampling vagaries as well as difficulties in timing sample collection. Therefore, in some systemic infections with elevated antibody responses, serodiagnosis may be preferable to isolation. In some infants we have recovered the organism from lung biopsies at a time when we could not culture it from the upper respiratory tract, but all these infants had high IgM antibody titers, which would have supported the diagnosis and removed the necessity for an invasive procedure.

CLINICAL SPECTRUM OF CHLAMYDIAL INFECTIONS

Male Genital Tract

Chlamydiae are the leading identified cause of nongonococcal urethritis (NGU) in men (2, 28). Most studies have shown that *C. trachomatis* can be associated with 25–50% of NGU. The etiology of the remainder of NGU cases is still obscure, although *Ureaplasma urealyticum* appears to be the best candidate for the majority of other cases (21, 29). Both *Chlamydia* and *Ureaplasma* are sensitive to tetracyclines and these are the drugs of choice in managing NGU (2, 30).

Routine epidemiologic treatment of female sex partners of men with NGU is indicated. One third of women whose male partners have NGU will have a chlamydial infection of the cervix and more than two thirds of those whose male partners are *Chlamydia*-positive will have such infections.

Most of the 20–35% of men with gonorrhea who have a concomitant infection with *C. trachomatis* will go on to develop postgonococcal urethritis (PGU), a specific subset of NGU (31–33). Chlamydial infections do not respond to the short courses of penicillins used to treat gonorrhea, so these men develop a urethritis despite eradication of the gonococcal infection. Because the gonococcus is also usually sensitive to tetracycline, the Center for Disease Control (CDC) recommends that tetracycline (2 g/day X 5 days) be considered as a first treatment choice (as are penicillin and ampicillin regimens) for gonorrhea. Treatment with tetracycline will presumably be effective against the incubating or established chlamydial infection and thus prevent the development of PGU. Since 2 g/day of tetracycline for 7 days is effective in treating NGU, a tetracycline regimen is a reasonable treatment of choice for all male urethritis.

There are some obvious problems with the tetracycline regimen. It will not eradicate incubating syphilis; there is always a compliance problem with any long-duration therapy; tetracyclines are contraindicated for treating pregnant women, one of the populations most needing antichlamydial therapy. To deal with some of these problems, Oriel suggests immediate treatment with penicillin for gonococcal infections, with the patient being given a follow-up of 7 days of oral tetracycline. Thus, the patient is immediately treated for the gonococcal infection and starts the tetracycline treatment the next day (removing the problem of combining a bacteriostatic and bacteriocidal drug).

Epididymitis appears to be the major complication of chlamydial urethritis in men. A study by Berger et al (35) identified *C. trachomatis* in epididymial aspirates from men with "idiopathic" epididymitis. *Chlamydia* were recovered from epididymitis at a higher frequency than gonococci in men younger than 35 years.

Chlamydia may also be associated with proctitis in either men or women (36, 37) but the prevalence of this infection and a causative association between *Chlamydia* and the rectal inflammation has not been proven. It is likely that *Chlamydia* contribute to the "gay bowel syndrome" and cause a nongonococcal proctitis just as they cause a nongonococcal urethritis.

Reiter's Syndrome will develop in some men following bouts of NGU. Some studies have found *C. trachomatis* in more than 60% of men with

Reiter's Syndrome and signs of urethritis (38). Since Reiter's Syndrome appears to occur in a genetically predisposed population (approximately 95% of patients are HLA B-27 positive) and the agent is rarely recovered from extragenital sites, the exact relationship of *Chlamydia* to this disease remains obscure; it has been speculated that *C. trachomatis* infection is a trigger mechanism (39).

Female Genital Tract

Holmes has pointed out that the clinical spectrum of chlamydial infections of the genital tract virtually parallels the spectrum of gonococcal genital tract infections (34). Thus, asymptomatic and clinically inapparent infections are recognized for both men and women. These are usually discovered during routine screening procedures or as a result of contact tracing from symptomatic patients. The site most commonly infected in women is the cervix. Chlamydial infection here appears to be restricted to the endocervical canal as the organism has never been shown to grow in vaginal squamous cells. Thus, it is not associated with vaginitis or ectocervicitis. The infected cervix may range from a clinically normal examination to a severely eroded cervix with hypertrophic cervical erosions and a mucopurulent endocervical discharge (40). Follicle-like lesions similar to those seen in the conjunctiva may develop in the cervix. Ectopy is often associated with *C. trachomatis,* but it can be induced by a number of other stimuli. In one large series, the only two organisms found to be associated with chronic cervicitis and purulent endocervical discharge were *C. trachomatis* and *N. gonorrhoeae* (41). It is possible that empirical management of cervicitis will evolve just as such management developed for NGU (i.e. rule out other pathogens and treat with tetracycline). Chlamydial infections of the cervix do respond to treatment with tetracycline (1 g/day X 14 days). Erythromycin is an alternative choice and is indicated for pregnant women.

C. trachomatis may also infect the urethra of women and such infection has been associated with dysuria (42). Of the *Chlamydia*-positive female partners of men with NGU, ~25% had the agent apparently restricted to the urethra, ~50% yielded organism from both cervix and urethra, and ~25% had infection restricted to the cervix. *C. trachomatis* is likely one of the causes of urethritis or the urethral syndrome in women.

The most serious complication of chlamydial infection in women is acute salpingitis. The rate at which this disease occurs is not known. In Sweden where the incidence of gonococcal infections has been decreasing, the rate of acute salpingitis has been increasing. In one study where isolation attempts were performed on material aspirated during laparoscopy from the involved fallopian tubes, *C. trachomatis* was recovered from 30% of the women with acute salpingitis (43). All evidence to date suggests that chlamydial etiology of acute salpingitis in the US is less, on the order of 10–20%. Comparable isolation data are not available since none of the research protocols being followed in the US involve needle biopsy of the fallopian tube (as is done in Sweden). Three studies that sought *C. trachomatis* in culdocentesis fluids or swabs at the fallopian tubes of women with acute salpingitis reported recovery rates ranging from 0 to 10% (44–46). Serologic evidence suggested that 20% of acute salpingitis in Seattle or San Francisco was associated with chlamydial infection (44, 46a). Studies from Scandinavia and England also note that some women with acute salpingitis have very high antibody titers of IgG or IgM class to *C. trachomatis* (22, 47). In addition, serologic evidence indicates an association between *C. trachomatis* and the Fitz-Hugh-Curtis Syndrome or the perihepatitis and peritonitis associated with some cases of salpingitis (48).

The means by which chlamydiae (or gonococci for that matter) gain access to the fallopian tubes is not known. Obviously, salpingitis usually represents an ascendence of a lower genital tract infection. It is not clear, however, whether the pathogens reach the tubes by attachment to spermatozoa, are spread by phagocytes, or move by extension of infection through the uterus. A characteristic common to both chlamydial and gonococcal infection in women is that both organisms cause an endocervicitis. It is tempting to speculate that this infection breaks a barrier protecting the uterus and fallopian tubes from vaginal flora. Thus, owing to the damage caused by chlamydial (or gonococcal) infection of the endocervix, flora from the lower genital tract might gain access to the fallopian tube. Because chlamydiae or gonococci are probably more invasive than most other organisms present in the vagina, they would be more likely to cause salpingitis. However, other organisms may be present in larger quantities and they may reach the tubes as a result of endocervical damage caused by chlamydiae and gonococci. This could explain the discordant results of cervical and tubal cultures that are seen in some cases of salpingitis. Intrauterine devices may also predispose to salpingitis by a similar mechanism—the tail may induce endocervical damage.

Infective genital tract discharges, when they reach the eye either by hand-to-eye transmission or during sexual activity, can cause an acute follicular conjunctivitis called inclusion conjunctivitis in adults (1, 3, 17). The cervix is also the source of infection to the newborn infant passing through the infected canal.

Infection in Neonates

CONJUNCTIVITIS The acute conjunctivitis of infants, inclusion conjunctivitis of the newborn (ICN) or inclusion blennorrhea, as it was termed in the early 1900s by the first workers to observe it, is a mucopurulent conjunctivitis that develops approximately 5–14 days after birth (17). Recent prospective studies have shown that approximately one third to one half of infants exposed at birth will develop ICN, presently the most common form of conjunctivitis in the first month of life (2, 24, 49–51). The disease is characterized by marked conjunctival erythema, moderate lymphoreticular proliferation, and copious purulent discharge. Untreated, it usually lasts several weeks, but eventually clears without affecting vision, although conjunctival scarring and micropannus may be seen. In severe cases diagnosis is readily made by demonstrating the typical inclusion bodies by a Giemsa stain of conjunctival scrapings (52). The *Chlamydia* are also readily cultured from the eye. Serologic diagnosis is not helpful because of the presence of maternally transmitted antichlamydial IgG antibody and the uncertain appearance of IgM in this illness (23, 24).

Time-honored treatment has been topical sulfonamides, erythromycin, or tetracycline ointments. Recent studies show that topical treatment alone is often inadequate for conjunctivitis. Furthermore, topical treatment does not affect nasopharyngeal carriage, nor does it prevent the development of respiratory tract disease (53, 54). For these reasons we recommend systemic treatment with erythromycin (40 mg/kg/day for 10–14 days).

CHLAMYDIAL PNEUMONIA The first suspected case of chlamydial pneumonia was observed in a prospective study aimed at defining the incidence of ICN (55). Shortly after this first report Beem & Saxon (23) published a series of retrospective and prospective cases of chlamydial pneumonia in young infants. Their report was followed by studies from other centers, and the clinical entity of chlamydial pneumonia is now well defined (56). While early studies included cases with concomitant viral infections (23, 57), it has become clear that *Chlamydia* can be and often are the sole causative agents of pneumonia (24, 58). This was shown by open lung biopsies on three infants with pneumonia and in several prospective studies. The pathology is one of mixed cell infiltration of the lung with obliteration of alveolar spaces and a severe necrotic bronchiolitis (58). An infant baboon model has been demonstrated and work is currently in progress on small-animal models of the disease (59).

Infants with this disease range from 2 weeks to 4 months old. Older children (up to 1 year of age) have been observed, but on careful history the pneumonia usually has had its onset before the fourth month of life. The disease may begin with rhinitis or mild cough. The classical picture described by Beem & Saxon is one of an afebrile, tachypneic infant with a prominent staccato cough, which occurs in paroxysms and has been called "pertussoid" (23, 53, 60). The characteristic whoop of pertussis is absent but apneic spells may occur. The lungs are overexpanded and the liver and spleen are easily felt because of the depression of the diaphragm. Some infants may have serious compromise of ventilation as evidenced by cyanosis and reduced pO_2 levels, but the majority of infants do not require oxygen or ventilatory support; pCO_2 may be elevated. A relative eosinophilia ($>$300 mm^3) is common. Chest film usually shows an interstitial diffuse pattern of pneumonia with some hyperinflation of the lungs.

Untreated, the disease may persist for weeks or months but eventually subsides without complications. Definitive diagnosis depends on either isolating chlamydiae from the lung or demonstrating a high level of specific IgM antibodies against *C. trachomatis* ($>$ 1:64). Demonstration of *C. trachomatis* in the nasopharynx or from aspiration of the tracheobronchial secretions is usually accepted as conclusive evidence of chlamydial etiology. However, this evidence alone does leave a small possibility that the chlamydiae were only infecting the respiratory passages while the pneumonia was viral in origin. We favor serological diagnosis because it is definitive, does not require invasive procedures like lung biopsies and deep tracheal suction, and in our experience has been more reliable than nasopharyngeal cultures alone. As a matter of practice we perform both nasopharyngeal cultures and serology. Differential diagnosis includes principally viral causes of pneumonia. Among these, respiratory syncytial virus and adenovirus are the most important. Cytomegalovirus has often been recovered in conjunction with chlamydial pneumonia (23, 57).

Recommended treatment is oral erythromycin (40 mg/kg/day in divided doses) for two weeks. Beem demonstrated that, while there is no dramatic response to therapy, treated infants recover more rapidly; furthermore, chlamydial isolation attempts become negative within 48 hours after initiation of such treatment (53, 60).

Prospective studies suggest that chlamydiae are probably responsible for some cases of bronchiolitis, and probably otitis media as well (chlamydiae have been recovered from tympanocentesis fluids) (60). Vaginal and rectal infections have been documented (61). The exact role of chlamydiae in these clinical syndromes or in producing disease at these sites remains to be elucidated.

Many of the early and continuing studies of chlamydial pneumonia in infants have demonstrated the frequent coexistence of chlamydiae and viral agents in the lung and respiratory passages. This is probably more than mere coincidence but reflects a high-risk group of infants most likely to acquire a variety of infectious agents early in life.

Harrison and colleagues, in a study of serial admissions for pneumonia, showed that *C. trachomatis* could be implicated in the etiology of 31% of consecutive pneumonia admissions in infants less than 6 months old (56).

Table 1 Prospective studies of perinatal chlamydial infection

Prevalence maternal infection(%)	Attack rate (%)		Cases per 1000 live births		Reference
	Conjunctivitis	Pneumonia	Conjunctivitis	Pneumonia	
12.7	50	not studied	63	not studied	51
8.8	44	11	40	10	50
5.0	40	not studied	20	not studied	2
4.0	35	20	14	8	24
2.0	33	16	7	3	49

In similar unpublished studies we have found *C. trachomatis* associated in different populations with between 15–50% of pneumonias in the first year of life. Thus, chlamydial infection is responsible for a major proportion of pneumonias in this age group. A number of prospective studies have been performed by testing pregnant women for chlamydial infection of the cervix and then following infants born through infected cervices (2, 24, 49–51). There is remarkable concordance in the results of these studies (Table 1): ~33–50% of exposed infants will develop conjunctivitis; between 10 and 20% of exposed infants develop pneumonia. Serologic evidence of infection is obtained in 60–70% of infants passing through the infected cervix. The incidence of perinatal chlamydial infection in the various study populations was a direct function of prevalence of chlamydial infection in the cervix, and given the constancy of the attack rates can probably be projected into other populations on that basis. Cervical chlamydial infection rates ranged from 2 to 12.7% in the studies that have been completely reported. Preliminary reports on other populations found infection rates as high as 25% (62). In all studies, the highest chlamydial infection rates in women were found in the highest-risk VD populations, i.e. those women who are unwed, young, and black (24, 50, 51).

Because the first cases of chlamydial pneumonia in infants appeared to be regularly associated with conjunctivitis, it was speculated that the eye might be the portal of entry for the organism, and pneumonia represented a seeding of the respiratory tract by infective ocular discharges (63). Thus it seemed possible that modification of Credé prophylaxis to use an anti-chlamydial ointment such as tetracycline or erythromycin might prevent ocular chlamydial infection and thus, prevent the subsequent development of pneumonia. Prospective studies, however, showed that most of the pneumonia cases were not preceded by conjunctivitis; the respiratory tract could be infected directly. A preliminary report has shown that neonatal prophylaxis with erythromycin ointment could prevent the development of ICN but had no apparent effect on nasopharyngeal infection or pneumonia

(64). Since chlamydial ICN is more common than gonococcal ophthalmia neonatorum and chlamydial infection rates of the cervix are higher than gonococcal infection rates of the cervix, it would seem reasonable to recommend routine prophylactic use of erythromycin ointment. This should be effective against gonococcal infections and would not be associated with the chemical conjunctivitis that invariably follows the silver nitrate instillation, which has no effect on chlamydiae.

REMAINING PROBLEMS AND CHALLENGES

C. trachomatis is clearly one of the major genital tract pathogens in our country. It is likely that the 3 million cases of *C. trachomatis* infection estimated to occur each year cost the US consumers more than 1 billion dollars a year in direct health costs (4). *C. trachomatis* is probably responsible for half of the 500,000 physician visits each year for epididymitis and for 20% of the 1 million visits for salpingitis.

Given the morbidity associated with *C. trachomatis* infections, some efforts at control measures are indicated. Unfortunately, effective control would require the introduction of chlamydial isolation procedures on a much broader scale than currently appears feasible. The first control measures would likely be aimed at venereal disease clinic populations. Since chlamydial infection rates are much higher than gonococcal infection rates outside VD clinics (65), this approach would miss a major proportion of the chlamydial infections. Chlamydial technology has been introduced into some medical facilities for the specific purpose of determining chlamydial infections in men with nongonococcal urethritis. This is the population least requiring such technology because the management of these patients would call for tetracycline treatment whether or not they had a proven chlamydial infection. It is far more important to introduce chlamydial diagnosis in managing genital tract complaints in women. They are at the greatest risk of developing serious complications of the infection and, of course, are epidemiologically crucial in both vertical and horizontal transmission. It is pregnant women who should be the first focus of chalmydial control procedures since each year in the US 155,000 infants are exposed to chlamydial infection and more than 100,000 will acquire these infections, with approximately 75,000 cases of conjunctivitis and 30,000 cases of pneumonia each year (4). In addition, these women are at risk for developing complications of pregnancy or postpartum salpingitis.

Without considering these maternal complications, it can be shown that screening pregnant women for chlamydial infection in an effort to prevent

Table 2 Cost-benefit analysis[a]

	Cost in dollars at prevalence of infection of				
	5%	10%	15%	20%	25%
Isolation tests per 1,000 women	$25,000	$ 25,000	$ 25,000	$ 25,000	$ 25,000
Follow-up and treating infections[b]	2,500	5,000	7,500	10,000	12,500
Cost of screening programs	27,500	30,000	32,500	35,000	37,500
Treating conjunctivitis[c]	1,250	2,500	3,750	5,000	6,250
Treating pneumonia[d]	25,000	50,000	75,000	100,000	125,000
Treating chlamydial disease	26,250	52,500	78,750	105,000	131,250
Dollars saved per 1,000 women screened	−1,250	+22,500	+46,250	+70,000	+93,750

[a] For screening pregnant women for chlamydial infection and preventing disease in their offspring.

[b] $50 per infected woman for treatment, follow-up visit, and test of cure.

[c] $50 per case for clinic visits and drugs; attack rate = 50%.

[d] $1,250 per case for hospitalization; average 4 days, drugs, laboratory tests; attack rate = 20%.

perinatal infections could be cost effective (Table 2). The nationwide prevalence of chlamydial infection of the cervix is generally assumed to be ~5%. At this rate, costs exceed benefits as the break-even point is ~6% prevalence. Chlamydial infections of the cervix have been found at > 6% levels in many studies of inner city populations (some have been > 20%, at which point financial savings would be considerable). Thus, it is specifically recommended that chlamydial isolation procedures should be introduced for routine screening of pregnant women in selected populations. The *Chlamydia*-positive women should be treated (presumably with erythromycin, although an appropriate regimen must be validated).

Isolation procedures are not technically difficult and could be established in the virology laboratories often associated with these clinics. Serology (micro-IF) is more difficult, but it should be made available, at least on a regional basis, to provide diagnostic assistance for chlamydial pneumonia in infants.

We hope the day will soon approach when appropriate technology will be available to accurately diagnose and treat these infections. However, even today a number of measures could be undertaken that should provide some modicum of control. Following CDC recommendations and using tetracycline to treat gonorrhea should reduce some of the chlamydial infections and postgonococcal urethritis. Routine use of tetracycline in treating men with NGU and in epidemiologic treatment of their sexual partners should reduce some of the infective pool. A similar approach could be applied to hypertrophic cervical erosion and mucopurulent endocervicitis in women. Obviously, chlamydial disease in an infant is an indicator of an infected genital tract of the mother and she and her sexual partner(s) should be treated. Since all available data indicate that nongonococcal urethritis is still increasing and is occurring at epidemic rates, and by inference chlamydial infections are also increasing, these measures should at least slow down the increase and provide some stopgap approach until more specific, directed procedures can be followed.

Literature Cited

1. Jones, B. R. 1974. Laboratory tests for chlamydial infection: Their role in epidemiological studies of trachoma and its control. *Br. J. Ophthalmol.* 58:438–54
2. Schachter, J. 1978. Chlamydial infections. *N. Engl. J. Med.* 298:428–35, 490–95, 540–49
3. Grayston, J. T., Wang, S-P. 1975. New knowledge of chlamydiae and the diseases they cause. *J. Infect. Dis.* 132:87–105
4. National Institute of Allergy and Infectious Diseases. 1980. *Chlamydia* and nongonococcal urethritis. *Sexually Transmitted Diseases Study Group Meeting.* In press
5. Page, L. A. 1974. The rickettsias. In *Bergey's Manual of Determinative Bacteriology,* ed. R. E. Buchanan, N. E. Gibbons, pp. 914–28. Baltimore: Williams & Wilkins. 8th ed.
6. Byrne, G. I., Moulder, J. W. 1978. Parasite-specified phagocytosis of *Chlamydia psittaci* and *Chlamydia trachomatis* by L and HeLa cells. *Infect. Immun.* 19(2):598–606
7. Kuo, C.-C., Wang, S.-P., Grayston, J. T. 1973. Effect of polycations, polyanions and neuraminidase on the infectivity of trachoma-inclusion conjunctivitis and lymphogranuloma venereum organisms in HeLa cells: Sialic acid residues as possible receptors for trachoma-inclusion conjunctivitis. *Infect. Immun.* 8:74–79
8. Levy, N. J. 1979. Wheat germ agglutinin blockage of chlamydial attachment sites: antagonism by *N*-acetyl-D-glucosamine. *Infect. Immun.* 25(3): 946–53
9. Friis, R. R. 1972. Interaction of L cells and *Chlamydia psittaci:* entry of the parasite and host responses to its development. *J. Bacteriol.* 110:706–21
10. Moulder, J. W. 1966. The relation of the psittacosis group (chlamydiae) to bacteria and viruses. *Ann. Rev. Microbiol.* 20:107–30
11. Schachter, J., Dawson, C. R. 1979. Psittacosis-lymphogranuloma venereum agents/TRIC agents. In *Diagnostic Procedures for Viral and Rickettsial Infections,* ed. E. H. Lennette, N. Schmidt, pp. 1021–59. Washington, DC: Am. Publ. Health Assoc. 5th ed.
12. Schachter, J. 1980. Chlamydiae. In *Manual of Clinical Immunology,* ed. F. H. Top, Washington, DC: Am. Soc. for Microbiol. 2nd ed. In press
13. Hanna, L., Schmidt, L., Sharp, M., Stites, D., Jawetz, E. 1979. Human cell-mediated immune response to chlamydial antigens. *Infect. Immun.* 23(2):412–17
14. Wang, S.-P., Grayston, J. T. 1970. Immunologic relationship between genital TRIC, lymphogranuloma venereum and related organisms in a new microtiter indirect immunofluorescence test. *Am. J. Ophthalmol.* 70:367–74
15. Schachter, J., Cles, L., Ray, R., Hines, P. 1979. Failure of serology in diagnosing chlamydial infections of the female genital tract. *J. Clin. Microbiol.* 10(5): 647–49
16. Richmond, S. J., Milne, J. D., Hilton, A. L., Caul, E. O. 1980. Antibodies to *Chlamydia trachomatis* in cervicovaginal secretions: relation to serum antibodies and current chlamydial infection. *Sex. Transm. Dis.* 7(1):11–15
17. Schachter, J., Dawson, C. R. 1978. *Human Chlamydial Infections.* Littleton, Mass: Publishing Sciences Group. 273 pp.

18. Rank, R. G., White, H. J., Barron, A. L. 1979. Humoral immunity in the resolution of genital infection in female guinea pigs infected with the agent of guinea pig inclusion conjunctivitis. *Infect. Immun.* 26(2):573–79
19. Ormsby, H. L., Thompson, G. A., Cousineau, G. G., Lloyd, L. A., Hassard, J. 1952. Topical therapy in inclusion conjunctivitis. *Am. J. Ophthalmol.* 35:1811–14
20. Tack, K. J., Rasp, F. L., Hanto, D., Peterson, P. K. O'Leary, M., Simmons, R. L., Sabath, L. D. 1980. Isolation of *Chlamydia trachomatis* from the lower respiratory tract of adults. *Lancet* 1:116–20
21. Bowie, W. R., Wang, S.-P., Alexander, E. R., Floyd, J., Forsyth, P. S., Pollock, H. M., Lin, J. S., Buchanan, T. M., Holmes, K. K. 1977. Etiology of nongonococcal urethritis: evidence for *Chlamydia trachomatis* and *Ureaplasma urealyticum. J. Clin. Invest.* 59:735–42
22. Treharne, J. D., Ripa, K. T., Mardh, P.-A., Svensson, L., Westrom, L., Darougar, S. 1979. Antibodies to *Chlamydia trachomatis* in acute salpingitis. *Br. J. Vener. Dis.* 55(1):26–29
23. Beem, M. O., Saxon, E. M. 1977. Respiratory-tract colonization and a distinctive pneumonia syndrome in infants infected with *Chlamydia trachomatis. N. Engl. J. Med.* 296:306–10
24. Schachter, J., Holt, J., Goodner, E., Grossman, M., Sweet, R., Mills, J. 1979. Prospective study of chlamydial infection in neonates. *Lancet* 2(8139): 377–80
25. Wentworth, B. B., Alexander, E. R. 1974. Isolation of *Chlamydia trachomatis* by use of 5-iodo-2-deoxyuridine-treated cells. *Appl. Microbiol.* 27: 912–16
26. Ripa, K. T., Mardh, P.-A. 1977. Cultivation of *Chlamydia trachomatis* in cycloheximide-treated McCoy cells. *J. Clin. Microbiol.* 6(4):328–31
27. Weiss, E., Dressler, H. R. 1960. Centrifugation of rickettsiae and viruses onto cells and on infection. *Proc. Soc. Exp. Biol. NY* 103:691–95
28. Holmes, K. K., Handsfield, H. H., Wang, S.-P., Wentworth, B. B., Turck, M., Anderson, J. B., Alexander, E. R. 1975. Etiology of nongonococcal urethritis. *N. Engl. J. Med.* 292:1199–1206
29. Bowie, W. R., Alexander, E. R., Floyd, J. F., Holmes, J., Miller, Y., Holmes, K. K. 1976. Differential response of chlamydial and ureaplasma-associated urethritis to sulphafurazole (sulfisoxazole) and aminocyclitols. *Lancet* 2:1276–78
30. Richmond, S. J., Oriel, J. D. 1978. Recognition and management of genital chlamydial infection. *Br. Med. J.* 2(6135):480–82
31. Richmond, S. J., Hilton, A. L., Clarke, S. K. R. 1972. Chlamydial infection: Role of *Chlamydia* subgroup A in nongonococcal and postgonococcal urethritis: *Br. J. Vener. Dis.* 48:437–44
32. Oriel, J. D., Reeve, P., Thomas, B. J., Nicol, C. S. 1975. Infection with *Chlamydia* group A in men with urethritis due to *Neisseria gonorrhoeae. J. Infect. Dis.* 131:376–82
33. Oriel, J. D., Ridgway, G. L., Reeve, P., Beckingham, D. C., Owen, J., 1976. The lack of effect of ampicillin plus probenecid given for genital infections with *Neisseria gonorrhoeae* on associated infections with *Chlamydia trachomatis. J. Infect. Dis.* 133(5):568–71
34. Schachter, J., Oriel, J. D., Dawson, C. R. Holmes, K. K., Beem, M. O. 1980. Human infection with *Chlamydia trachomatis.* In *Current Chemotherapy and Infectious Disease: Proc. 11th Int. Congr. Chemother., 19th Intersci. Conf. on Antimicrobial Agents Chemother.*, ed. J. D. Nelson, C. Grassi, 1:31–35. Washington, DC: Am. Soc. for Microbiol.
35. Berger, R. E., Alexander, E. R., Monda, G. D., Ansell, J., McCormick, G., Holmes, K. K. 1978. *Chlamydia trachomatis* as a cause of acute "idiopathic" epididymitis. *N. Engl. J. Med.* 298(6):301–4
36. Dunlop, E. M. C., Harper, I. A., Al-Hussaini, M. K., Garland, J. A., Treharne, J. D., Wright, D. J. M., Jones, B. R. 1966. Relation of TRIC agent to "non-specific" genital infection. *Br. J. Vener. Dis.* 42:33–42
37. Goldmeier, D., Darougar, S. 1977. Isolation of *Chlamydia trachomatis* from throat and rectum of homosexual men. *Br. J. Vener. Dis.* 53:184–85
38. Kousa, M., Saikku, P., Richmond, S. J., Lassus, A. 1978. Frequent association of chlamydial infection with Reiter's Syndrome. *Sex. Transm. Dis.* 5(2): 57–61
39. Schachter, J. 1976. Can chlamydial infections cause rheumatic disease? In *Infection and Immunology in the Rheumatic Diseases,* ed. D. C. Dumonde, pp. 151–57. Oxford: Blackwell Scientific
40. Oriel, J. D., Johnson, A. L., Barlow, D., Thomas, B. J., Nayyar, D., Reeve, P.

1978. Infection of the uterine cervix with *Chlamydia trachomatis*. *J. Infect. Dis.* 137(4):443–51

41. Rees, E., Tait, I. A., Hobson, D., Johnson, R. W. A. 1977. *Chlamydia* in relation to cervical infection and pelvic inflammatory disease. In *Nongonococcal Urethritis and Related Infections*, ed. K. K. Holmes, D. Hobson, pp. 67–76. Washington, DC: Am. Soc. for Microbiol.
42. Paavonen, J. 1979. Chlamydia trachomatis-induced urethritis in female partners of men with nongonococcal urethritis. *Sex. Transm. Dis.* 6(2):69–71
43. Mardh, P.-A., Ripa, K. T., Svensson, L., Westrom, L. 1977. *Chlamydia trachomatis* infection in patients with acute salpingitis. *N. Engl. J. Med.* 296:1377–79
44. Eschenbach, D. A., Buchanan, T. M., Pollock, H. M., Forsyth, P. S., Alexander, E. R., Lin, J.-S., Wang, S.-P., Wentworth, B. B., McCormack, W. M., Holmes, K. K. 1975. Polymicrobial etiology of acute pelvic inflammatory disease. *N. Engl. J. Med.* 293:166–71
45. Thompson, S. E., Hager, W. D., Wong, K.-H., Lopez, B., Ramsey, C., Allen, S. D., Stargel, M. D., Thornsberry, C., Benigno, B. B., Thompson, J. D., Shulman, J. A. 1980. The microbiology and therapy of acute pelvic inflammatory disease in hospitalized patients. *Am. J. Obstet. Gynecol.* 136:179–86
46. Sweet, R. L., Mills, J., Hadley, K. W., Blumenstock, E., Schachter, J., Robbie, M. O., Draper, D. L. 1979. Use of laparoscopy to determine the microbiologic etiology of acute salpingitis. *Am. J. Obstet. Gynecol.* 134(1):68–74

46a. Sweet, R. L., Draper, D. L., Schachter, J., James, J., Hadley, W. K., Brooks, G. F. 1980. Microbiology and pathogenesis of acute salpingitis as determined by laparoscopy: what is the appropriate site to sample. Presented at the Int. Symp. on Pelvic Inflammatory Dis., Atlanta, Ga. *Am. J. Obstet. Gynecol.* In press

47. Paavonen, J., Saikku, P., Vesterinen, E., Aho, K. 1979. *Chlamydia trachomatis* in acute salpingitis. *Br. J. Vener. Dis.* 55(3):203–6
48. Muller-Schoop, J. W., Wang, S.-P., Munzinger, J., Schlapfer, H. U., Knoblauch, M., Ammann, R. W. 1978. *Chlamydia trachomatis* as possible cause of peritonitis and perihepatitis in young women. *Br. Med. J.* 1(6119): 1022–23
49. Hammerschlag, M. R., Anderka, M., Semine, D. Z., McComb, D., McCormack, W. M. 1979. Prospective study of maternal and infantile infection with *Chlamydia trachomatis*. *Pediatrics* 64(2):142–48
50. Frommell, G. T., Rothenberg, R., Wang, S.-P., McIntosh, K., Wintersgill, C., Allaman, J., Orr, I. 1979. Chlamydial infection of mothers and their infants. *J. Pediatr.* 95(1):28–32
51. Chandler, J. W., Alexander, E. R., Pheiffer, T. A., Wang, S.-P., Holmes, K. K., English, M. 1977. Ophthalmia neonatorum associated with maternal chlamydial infections. *Trans. Am. Acad. Ophthalmol. Otolaryn.* 83:302–8
52. Schachter, J., Dawson, C. R. 1977. Comparative efficiency of various diagnostic methods for chlamydial infection. See Ref. 41, pp. 337–41
53. Beem, M. O., Saxon, E. M., Tipple, M. 1979. Treatment of chlamydial pneumonia in infancy. *Pediatrics* 63(2):198–203
54. Rowe, S., Aicardi, E., Dawson, C. R., Schachter, J. 1979. Purulent ocular discharge in neonates: significance of *Chlamydia trachomatis*. *Pediatr. Pediau.* 63(4):628–32
55. Schachter, J., Lum, L., Gooding, C. A., Ostler, B. 1975. Pneumonitis following inclusion blennorrhea. *J. Pediatr.* 87(5): 779–80
56. Harrison, H. R., English, M. G., Lee, C. K., Alexander, E. R. 1978. *Chlamydia trachomatis* infant pneumonitis: Comparison with matched controls and other infant pneumonitis. *N. Engl. J. Med.* 298(13):702–8
57. Frommell, G. T., Bruhn, F. W., Schwartzman, J. D. 1977. Isolation of *Chlamydia trachomatis* from infant lung tissue. *N. Engl. J. Med.* 296: 1150–52
58. Arth, C., Von Schmidt, B., Grossman, M., Schachter, J. 1978. Chlamydial pneumonitis. *J. Pediatr.* 93(3):447–49
59. Harrison, H. R., Alexander, E. R., Chiang, W.-T., Giddens, W., Boyce, J., Benjamin, D., Gale, J. 1979. Experimental nasopharyngitis and pneumonia caused by *Chlamydia trachomatis* in infant baboons: histopathologic comparison with a case in a human infant. *J. Infect. Dis.* 139(2):141–46
60. Tipple, M., Beem, M. O., Saxon, E. 1979. Clinical characteristics of the afebrile pneumonia associated with *Chlamydia trachomatis* infection in infants less than 6 months of age. *Pediatrics* 63(2):192–97

61. Schachter, J., Grossman, M., Holt, J., Sweet, R., Spector, S. 1979. Infection with *Chlamydia trachomatis:* involvement of multiple sites in neonates. *J. Infect. Dis.* 139(2):232–34
62. Lumicoa, G. G., Gyves, M. T., Stuart, L. A., Heggie, A. D. 1979. Prospective study of perinatal infection with *Chlamydia trachomatis. J. Pediatr. Res.* 13:464 (Abstr. 829).
63. Schachter, J. 1977. The expanding clinical spectrum of infections with *Chlamydia trachomatis. Sex. Transm. Dis.* 4(3):116–18
64. Hammerschlag, M. R., Chiang, W. T., Chandler, J. W., English, M., Alexander, E. R. 1979. Effectiveness of erythromycin neonatal ocular prophylaxis in prevention of infant *Chlamydia trachomatis* infection. *Programs and Abstracts: 11th Int. Congr. Chemother., 19th Intersci. Conf. on Antimicrobial Agents and Chemother.* Washington, DC: Am. Soc. for Microbiol. (Abstr. 524)
65. Schachter, J., Hanna, L., Hill, E. C., Massad, S., Sheppard, C. W., Conte, J. E. Jr., Cohen, S. N., Meyer, K. F. 1975. Are chlamydial infections the most prevalent venereal disease? *J. Am. Med. Assoc.* 231(2):1252–55

Ann. Rev. Med. 1981. 32:63–71
Copyright © 1981 by Annual Reviews Inc. All rights reserved

CHEMICAL HEADACHES

❖7382

Neil H. Raskin, M.D.

Department of Neurology, University of California School of Medicine,
San Francisco, California 94143

It is not entirely clear which types of headache originate centrally and/or peripherally. The most common type of pain is that which results from activation of peripheral receptors in the presence of a normally functioning nervous system, as in the pain arising from peptic ulcer or subungual hematoma. Pain may also arise as the result of injury to the central or peripheral nervous system and may occur without activation of peripheral receptors (1). Headache may arise from dysfunction of, displacement of, or encroachment on pain-sensitive cranial structures that, for the most part, are vascular (2). The traditional view of the origin of chemically induced headaches is that receptors in arterial walls are stimulated or inhibited, which thereby causes vasodilatation and results in the perception of pain. However, it has recently become evident that the capacity for vasomotor regulation resides within the central nervous system itself (3, 4), so that in addition to direct effects on vascular smooth muscle, chemicals may modify vascular reactivity via brainstem mechanisms. This latter possibility appears to be more consistent with the diverse peripheral pharmacologic properties of the chemicals that may provoke headache.

Clinically, chemical headaches exhibit certain "vascular" features which support the hypothesis that dilatation of intra- and/or extracranial arteries or erratic modulation of vasomotor tone is important to their mechanism. They are often pulsatile in quality when severe and are often intensified by motion of the head, physical exertion, coughing or straining, or a change in posture. Occasionally, transient flashes of light or star figures also accompany vascular headaches (5).

0066-4219/81/0401-0063$01.00

NITROGLYCERINE

The first of the chemical headaches to be well characterized was the "dynamite headache" that occurred at the turn of the century among factory workers in plants manufacturing explosives (6). Nitroglycerine (NTG) is a fundamental component of dynamite and was eventually implicated as the offending agent (7); headache attends the administration of 0.3 or 0.4 mg of the drug when used for angina pectoris in over 50% of patients (8).

Transient monocular or binocular amaurosis occasionally precedes headache, which usually appears within minutes to an hour or more after exposure to NTG. The headache is usually located bifrontally or bitemporally, is almost invariably described as pulsatile in quality, and is markedly worsened by physical exertion or bending forward (9). Clinical doses of NTG result in headache episodes that seldom last more than 20 minutes, but occasionally attacks will last for several hours. A dose of NTG insufficient to cause headache may be augmented within minutes by the ingestion of only modest amounts of alcoholic beverages (10).

Workers in munitions plants soon acquire a tolerance to NTG, but it can be rapidly lost over a weekend. To prevent Monday morning headaches, workers have been known to rub NTG into their skin or use impregnated hat sweat bands on weekends (11).

Migraineurs probably have a lower threshold to NTG-provoked headache than do non-headache-prone individuals (12); serotonin metabolites appear in the urine of migraineurs after a NTG-provoked headache attack (13). Whereas ergotamine may abort 40% of NTG-provoked headaches (among non-migraineurs) (7), pretreatment with propranolol in daily doses ranging from 80 to 160 mg has not altered the headache response to NTG (8).

The peak vasodilator effects of NTG peripherally (14) and centrally (15) occur within three to four minutes of its administration and disappear in approximately 30 minutes. The occurrence of monocular visual loss preceding headache in some patients and the occasional lag of an hour or more before headache occurs suggest that factors other than, or in addition to, vasodilatation must be held accountable.

NITRITES

Many people experience varying degrees of headache minutes to hours after eating frankfurters or other cured meat products ("hot dog" headache) (16). The evidence implicates the nitrite content of these foodstuffs as the cause of such headaches. Similar to NTG-induced headaches, they are usually bifrontal or bitemporal, are of pulsatile quality about half the time, and are sometimes accompanied by facial flushing. These patients are sensitive to as little as one milligram of sodium nitrite, a dose that is hardly likely to exert direct effects upon arterial walls (17).

The observation that nitrate impurities in rock salt caused red patches in cured meat led to the deliberate use of fixed concentrations of salt and nitrate to produce a more uniformly colored product. Eventually it became clear that the coloring agent was not sodium nitrate, but sodium nitrite, which is formed in meat by the bacterial and chemical reduction of nitrate; nitrites have been generally substituted for nitrates in the curing process. Nitrite and its decomposition product, nitric oxide, react with myoglobin and hemoglobin to form the red compounds nitrosomyoglobin and nitrosohemoglobin (18). Government regulations limit the nitrite levels reached during the treatment of cured meats to 200 mg per kg of meat. Cooking and storage lead to a reduction in the nitrite concentration, so that the nitrite content of cured meat is approximately 50–130 mg per kg.

MONOSODIUM GLUTAMATE

About 30% of people who eat Chinese food report adverse reactions (19). Headache and tightness around the face are the most commonly reported symptoms, and many also report dizziness, diarrhea, nausea, and abdominal cramps; these symptoms constitute the "Chinese restaurant syndrome." Monosodium glutamate (MSG), the chemical precipitant of these symptoms (20), is one of the active ingredients in soy sauce and is used as a food additive for its flavor-enhancing properties, more so by some Chinese chefs than others. MSG injected intravenously produces a stereotyped sequence of symptoms that begins with a burning sensation of the chest which then spreads to the neck, shoulder, upper limbs, and abdomen. This is followed by a sensation of pressure over the chest, and finally by tightness and pressure over the face.

Orally administered MSG produces symptoms in about 20 minutes. The oral dosage threshold for minimal symptoms is from 1.5 to 12 gm. Oral MSG thresholds of 3 gm or less are usually found in subjects who report symptoms after eating Chinese food. A 200-ml portion of wonton soup, a commonly implicated menu item, contains about 3 gm of MSG. The previous ingestion of food delays absorption of MSG so that susceptible individuals may be protected from experiencing symptoms. Plasma glutamate levels after glutamate ingestion are not different in MSG-sensitive and MSG-nonsensitive persons (21). Animals fed glutamate show unchanged brain glutamate levels but brain glutamine increases (22). Glutamate is present in high concentrations throughout the body and is synthesized in brain and retina;

how its ingestion produces symptoms is not clear. It has been suggested that glutamate-induced release of acetylcholine may explain the syndrome (23). MSG exerts no known actions upon arterial walls.

TYRAMINE

By the turn of the century, the amine derivative of tyrosine, tyramine, had been extracted from cheese and was found to cause hypertension and headache when taken orally (24) or parenterally (25). Subsequently, it was shown that the pressor effects of tyramine were mediated by the release of norepinephrine from its binding sites (26).

Interest in tyramine escalated during the early 1960s when it was implicated in a syndrome of intense throbbing headache, sometimes associated with hypertension, that occurred in some patients receiving monoamine oxidase inhibitors after they had eaten foods rich in tyramine (27). Hanington & Harper (28) observed that the foodstuffs implicated in these untoward reactions were similar to those foods often reported by migraine sufferers as precipitants of their attacks, a condition that has come to be known as "dietary migraine." This subpopulation of migraineurs is sensitive to tyramine. A migraine attack was provoked within hours after the ingestion of 125 mg of tyramine in over 80% of instances, whereas neither the non-dietary-migraine patients nor the headache-free control subjects reported headaches to any significant extent (29, 30).

These results were confirmed in part by Bonnet & Lepreux (31), who identified 63 tyramine-sensitive migraineurs among a series of 213 cases. Ghose and associates (32), using intravenous tyramine, precipitated headache attacks in 46% of 31 migraineurs and in none of 27 control subjects. On the other hand, Moffett and co-workers (33) and Shaw et al (34), in studies of eight and nine diet-sensitive patients with migraine respectively, were unable to demonstrate tyramine sensitivity. Hanington (35) responded to the failure of confirmation by pointing out the methodologic pitfalls that may account for the discrepant results. On balance, the evidence supporting tyramine sensitivity in dietary migraine is persuasive.

Sulfate conjugation forms an important inactivation mechanism for ingested monoamines (36), and two reports have suggested that this mechanism is defective in tyramine-sensitive migraineurs (37, 38). A subsequent study from one of these laboratories in which methods of greater precision were used failed to confirm the original observation (39).

Although endogenous tyramine is located primarily within the central nervous system (40), where it may function in the modulation of norepinephrine release, the mechanism by which tyramine is implicated in headache is not clear. It appears likely that exogenous tyramine can trigger headache, but blood pressure elevations do not occur at the doses used. Tyramine releases serotonin from platelets in vitro (41), but this phenomenon has not been correlated with headache responsiveness.

PHENYLETHYLAMINE

Although tyramine-containing dairy products are frequently implicated as headache precipitants in dietary migraine, other provocative foodstuffs such as ethyl alcohol and chocolate contain little or no tyramine (42). Another vasoactive amine, phenylethylamine (PEA), has been identified in a variety of chocolates, cheeses, and some red wines (43, 44); unsweetened chocolate contains the highest concentration of PEA, approximately 0.4 mg per ounce. Like tyramine, endogenous PEA is present in human brain and may modulate functions influenced by catecholamines (45); exogenous PEA crosses the blood-brain barrier and can effect large changes in cerebral blood flow (46). Using 3 mg of PEA, Sandler et al (47) showed that patients with chocolate-provoked migrainous attacks reported headache about 50% of the time after PEA ingestion, compared to a 15% incidence after placebo ingestion ($p<0.01$). Moffett et al (48), using 44–62-gm samples of two types of chocolate of unknown PEA content, were unable to provoke headache significantly more often than after placebo ingestion. On balance, the evidence implicating PEA in chocolate-induced headaches is persuasive.

ETHYL ALCOHOL

Alcoholic beverages rank second to chocolate in provoking headache in susceptible individuals (29, 48). Headache usually appears within 30–45 minutes after alcohol consumption, which corresponds to the time it takes for alcohol to dilate skin arterioles (49). However, alcohol has little or no effect on muscle or cerebral blood flow (15, 49), so vasodilatation is probably not alcohol's mode of action in provoking headache. That alcoholemia interferes with cerebral autoregulation (50) offers more promise in tracking down the mechanism of "cocktail headache."

The hangover syndrome appears to be a disparate disorder. In addition to headache, we find malaise, nausea, pallor, and hyper-excitability occurring in varying admixtures several hours after the intake of alcohol has ceased, at a time when tissue levels of alcohol are very low or zero. Headache is a common but not constant feature of the syndrome; it is usually generalized, commonly throbbing, and worsened by postural change, coughing, and rapid movement of the head. The syndrome usually lasts 5–10 hours, long after all the alcohol has been metabolized (51). Alterations of electrolyte and water balance correlate poorly with hangover symptoms (52). The immediate lessening of symptoms after drinking alcohol has led to the suggestion that the hangover state may be a mild withdrawal symptom.

Many social drinkers believe that there are differences in both the intoxication and the hangover that result from drinking different alcoholic beverages. Many beverages contain a great number of congeners in addition to ethanol and water; high congener beverages such as bourbon and cognac are more likely to result in hangover than low congener beverages such as vodka (53, 54). The most pharmacologically active congeners are the aldehydes, methanol, and the fusel alcohols. The minute levels of the aldehydes and methanol in alcoholic beverages render it unlikely that they are implicated in the sequelae of drinking alcohol. The fusel alcohols comprise various aliphatic and aromatic alcohols formed during fermentation, among which are isobutyl, isoamyl, amyl, and *n*-propyl alcohols. They are rapidly metabolized, and assessments of a variety of physiologic functions could attribute no more than modest effects to these compounds (55). How or whether these substances alter the metabolic consequences of alcohol oxidation has not been studied.

Ethyl alcohol is oxidized to acetyldehyde, which is subsequently oxidized to acetate. Acetate enters the two-carbon pool, most of which is eventually expired as carbon dioxide (56). Since each oxidation is catalyzed by a dehydrogenase for which nicotinamide adenine dinucleotide (NAD) is a cofactor, there is a net accumulation of reduced NAD in the liver as a result of ethanol oxidation, probably accounting for the mild lactic acidosis that accompanies alcohol ingestion (57). Delayed clearance of the metabolic products of ethanol oxidation has been suspected in the genesis of the hangover syndrome: Ogata et al (58) found elevated blood levels of pyruvate, but not of lactate or acetaldehyde, in subjects with hangover symptoms. The widespread observation made by pilots that the inhalation of oxygen effectively ameliorates hangover symptoms (N. Raskin, unpublished) is consistent with the hypothesis that delayed metabolic restitution of the altered redox state that accompanies alcohol ingestion bears on the mechanism of hangover.

Alcohol dehydrogenase, the catalyst for the initial, rate-limiting oxidation of ethyl alcohol, is present in cerebral tissue (59). Although contributing little to the overall metabolism of an alcohol load, variation in the oxidation of alcohol at cellular sites within the brain may have important effects on neuronal function, and bear on the mechanism of hangover. The depression of brain serotonin turnover by a high blood alcohol level (60) may also be relevant because of the implication of synaptic serotonin in migraine (5).

CONCLUSION

There is no evidence to support the idea that allergy has a role in chemical headaches (61). In general, people who experience headache after exposure to one or more chemicals are very likely to develop spontaneous unprovoked migraine headaches also (28). Some individuals have a lowered biological threshold to nitroglycerine, MSG, alcohol, tyramine, and probably many other chemicals, and this makes them susceptible to vascular headaches under certain conditions. These chemicals have variable actions on arterial walls; hence simple direct vasodilatation is an unlikely mode of action. Moreover, the diverse pharmacologic properties of these compounds suggests that there is a common secondary, humoral mediator or central locus of action. The latter is more consistent with other observations that have been made regarding the mechanism of nonchemical headaches (5); it is tantalizing to speculate that this putative central threshold may reside within brainstem serotonergic circuits.

Rapid modifications of brain serotonin levels occur after the ingestion of a variety of chemicals (62). For neurons of the central nervous system, the first step in the synthesis of serotonin is uptake from the plasma of tryptophan, which is derived primarily from the diet (63). An active uptake process facilitates the entry of tryptophan into brain so that other substances may compete with tryptophan for entry into neurons with a resultant lowering of the serotonin level. Several lines of evidence suggest that lowering of the serotonin concentration at synapses in the midbrain raphe region of the brainstem increases the firing rates of serotonergic neurons (64), which may then produce the circulatory alterations (4) that account for vascular headaches. Thus, it is possible that central serotonergic neuronal activity can be modulated indirectly by chemicals competing for tryptophan entry into brain; it is also possible that chemicals directly affect serotonergic neuronal activity. Both hypotheses can be tested in a controlled laboratory setting.

Literature Cited

1. Basbaum, A. I., Fields, H. L. 1978. Endogenous pain control mechanisms: review and hypothesis. *Ann. Neurol.* 4: 451–62
2. Ray, B. S., Wolff, H. G. 1940. Experimental studies on headache. Pain-sensitive structures of the head and their significance in headache. *Arch. Surg.* 41: 813–56
3. Purves, M. J. 1978. Do vasomotor nerves significantly regulate cerebral blood flow? *Circ. Res.* 43:485–93
4. Reinhard, J. F. Jr., Liebmann, J. E., Schlosberg, A. J., Moskowitz, M. A. 1979. Serotonin neurons project to small blood vessels in the brain. *Science* 206:85–87
5. Raskin, N. H., Appenzeller, O. 1980. *Headache. Major Problems in Internal Medicine,* Vol. 19. Philadelphia: Saunders. 244 pp.
6. Laws, G. C. 1898. The effects of nitroglycerine upon those who manufacture it. *J. Am. Med. Assoc.* 31:793–94

7. Schwartz, A. M. 1946. The cause, relief and prevention of headaches arising from contact with dynamite. *N. Engl. J. Med* 235:541–44
8. Horwitz, L. D., Herman, M. V., Gorlin, R. 1972. Clinical response to nitroglycerin as a diagnostic test for coronary artery disease. *Am. J. Cardiol.* 29: 149–53
9. Rabinowich, I. M. 1944. Acute nitroglycerine poisoning. *Can. Med. Assoc. J.* 50:199–202
10. Ebright, G. E. 1914. The effects of nitroglycerine on those engaged in its manufacture. *J. Am. Med. Assoc.* 62:201–2
11. Abrams, J. 1980. Nitrate tolerance and dependence. *Am. Heart J.* 99:113–23
12. Schnitker, M. T., Schnitker, M. A. 1947. A clinical test for migraine. *J. Am. Med. Assoc.* 135:89
13. Campus, S., Fabris, F., Rappelli, A., Gastaldi, L., Gai, V., Nattero, G. 1967. Escrezione urinaria di acido 5-indrossiindolacetico durante crisi cefalalgica indotta da trinitro glicerina. *Boll. Soc. Ital. Biol. Sper.* 43:1844–47
14. Mason, D. T., Braunwald, E. 1965. The effects of nitroglycerine and amyl nitrite on arteriolar and venous tone in the human forearm. *Circulation* 32:755–66
15. Sokoloff, L. 1959. The action of drugs on the cerebral circulation. *Pharmacol. Rev.* 11:1–85
16. Henderson, W. R., Raskin, N. H. 1972. "Hot dog" headache: individual susceptibility to nitrite. *Lancet* 2:1162–63
17. Crandall, L. A. Jr., Leake, C. D., Loevenhart, A. S., Muehlberger, C. W. 1931. Acquired tolerance to and cross tolerance between the nitrous and nitric acid esters and sodium nitrite in man. *J. Pharmacol. Exp. Ther.* 41:103–19
18. Halliday, D. 1967. Curing of pig meat. *Process. Biochem.* 2:32–34
19. Reif-Lehrer, L. 1977. A questionnaire study of the prevalence of Chinese restaurant syndrome. *Fed. Proc.* 36: 1617–23
20. Schaumberg, H. H., Byck, R., Gerstl, R., Mashman, J. H. 1969. Monosodium L-glutamate: its pharmacology and role in the Chinese restaurant syndrome. *Science* 163:826–28
21. Kenney, R. A., Tidball, C. S. 1972. Human susceptibility to oral monosodium L-glutamate. *Am. J. Clin. Nutr.* 25: 140–46
22. Reif-Lehrer, L. 1976. Possible significance of adverse reactions to glutamate in humans. *Fed. Proc.* 35:2205–12
23. Ghadimi, H., Kumar, S., Abaci, F. 1971. Studies on monosodium glutamate ingestion. *Biochem. Med.* 5:447–556
24. Dale, H. H., Dixon, W. E. 1909. Action of pressor bases produced by putrefaction. *J. Physiol.* 34:25–44
25. Findlay, J. 1911. The systolic pressure at different points of the circulation in the child and the adult. *Q. J. Med.* 4:489–497
26. Burn, J. H., Rand, M. J. 1958. The action of sympathomimetic amines in animals treated with reserpine. *J. Physiol.* 144:314–36
27. Blackwell, B., Marley, E., Price, J., Taylor, D. 1967. Hypertensive interactions between monoamine oxidase inhibitors and foodstuffs. *Br. J. Psychiatry* 113:349–65
28. Hanington, E., Harper, A. M. 1968. The role of tyramine in the aetiology of migraine, and related studies on the cerebral and extracerebral circulations. *Headache* 8:84–97
29. Smith, I., Kellow, A. H., Hanington, E. 1970. A clinical and biochemical correlation between tyramine and migraine headache. *Headache* 10:43–52
30. Hanington, E., Horn, M., Wilkinson, M. 1970. Further observations on the effect of tyramine. In *Background to Migraine, 3rd Migraine Symposium,* ed. J. N. Cumings, pp. 113–19. New York: Springer. 186 pp.
31. Bonnet, G. F., Lepreux, P. 1971. Les migraines tyraminiques. *Sem. Hop. Paris* 47:2441–45
32. Ghose, K., Coppen, A., Carroll, D. 1977. Intravenous tyramine response in migraine before and during treatment with indoramin. *Br. Med. J.* 1:1191–93
33. Moffett, A. M., Swash, M., Scott, D. F. 1972. Effect of tyramine in migraine: a double-blind study. *J. Neurol. Neurosurg. Psychiatry* 35:496–99
34. Shaw, S. W. J., Johnson, R. H., Keogh, H. J. 1978. Oral tyramine in dietary migraine sufferers. In *Current Concepts in Migraine Research,* ed. R. Greene, pp. 31–39. New York: Raven. 169 pp.
35. Hanington, E. 1974. Monoamine oxidase and migraine. *Lancet* 2:1148–49
36. Connolly, M. E., Davies, D. S., Dollery, C. T., Morgan, C. D., Paterson, J. W., Sandler, M. 1972. Metabolism of isoprenaline in dog and man. *Br. J. Pharmacol.* 46:458–72
37. Smith, I., Mullen, P. E., Hanington, E. 1971. Dietary migraine and tyramine metabolism. *Nature* 230:246–48

38. Youdim, M. B. H., Bonham-Carter, S., Sandler, M., Hanington, E., Wilkinson, M. 1971. Conjugation defect in tyramine-sensitive migraine. *Nature* 230: 127–28
39. Smith, I., Gordon, A. J., Hanington, E., Marsh, S. E., Wilkinson, M. I. P. 1973. Studies of tyramine metabolism in migraine patients and control subjects. In *Background to Migraine, 5th Migraine Symposium,* ed. J. N. Cumings, pp. 34–39. New York: Springer. 182 pp.
40. Spector, S., Melmon, K., Lovenberg, W., Sjoerdsma, A. 1963. The presence and distribution of tyramine in mammalian tissues. *J. Pharmacol. Exp. Ther.* 140:229–35
41. Bartholini, G., Pletscher, A. 1964. Two types of 5-hydroxy-tryptamine release from isolated blood platelets. *Experientia* 20:376–78
42. Sandler, M. 1975. Monoamines and migraine: a path through the wood? In *Vasoactive Substances Relevant to Migraine,* ed. S. Diamond, D. J. Dalessio, J. R. Graham, J. L. Medina, pp. 3–18, Springfield, Ill: Thomas. 102 pp.
43. Chaytor, J. P., Crathorne, B., Saxby, M. J. 1975. The identification and significance of 2-phenylethylamine in foods. *J. Sci. Food Agric.* 26:593–98
44. Schweitzer, J. W., Friedhoff, A. J., Schwartz, R. 1975. Chocolate, β-phenylethylamine and migraine reexamined. *Nature* 257:256
45. Sabelli, H. C., Borison, R. L., Diamond, B. J., Havdala, H. S. 1978. Phenylethylamine and brain function. *Biochem. Pharmacol.* 27:1729–30
46. McCulloch, J., Harper, A. M. 1979. Factors influencing the response of the cerebral circulation to phenylethylamine. *Neurology* 29:201–7
47. Sandler, M., Youdim, M. B. H., Hanington, E. 1974. A phenylethylamine oxidising defect in migraine. *Nature* 250:335–37
48. Moffett, A. M., Swash, M., Scott, D. F. 1974. Effect of chocolate in migraine: a double-blind study. *J. Neurol. Neurosurg. Psychiatry* 37:445–48
49. Gillespie, J. A. 1967. Vasodilator properties of alcohol. *Br. Med. J.* 2:274–77
50. McQueen, J. D., Sklar, F. K., Posey, J. B. 1978. Autoregulation of blood flow during alcohol infusion. *J. Stud. Alcohol* 39:1477–87
51. Goldberg, L. 1961. Alcohol, tranquilizers and hangover. *Q. J. Stud. Alcohol* (Suppl.) 1:37–58
52. Flynn, P. A. 1959. Water and electrolyte balance after alcohol ingestion. *Q. J. Stud. Alcohol* 20:128–29
53. Damrau, F., Liddy, E. 1960. Hangovers and whiskey congeners. *J. Natl. Med. Assoc.* 52:262–65
54. Murphree, A. B., Greenberg, L. A., Caroll, R. B. 1967. Neuropharmacological effects of substances other than ethanol in alcoholic beverages. *Fed. Proc.* 26:1468–73
55. Wallgren, H., Barry, H. 1970. Drug actions in relation to alcohol effects. In *Actions of Alcohol,* 2:621–714. New York: Elsevier. 871 pp.
56. Hawkins, R. D., Kalant, H. 1972. The metabolism of ethanol and its metabolic effects. *Pharmacol. Rev.* 24:67–157
57. Isselbacher, K. J. 1977. Metabolic and hepatic effects of alcohol. *N. Engl. J. Med.* 296:612–16
58. Ogata, S., Hosoi, T., Saji, H. 1966. Studies on acute alcohol intoxication. *Jpn. J. Stud. Alcohol* 1:67–79
59. Raskin, N. H., Sokoloff, L. 1968. Brain alcohol dehydrogenase. *Science* 162: 131–32
60. Myers, R. D., Melchior, C. L. 1978. Alcohol and alcoholism: role of serotonin. In *Serotonin in Health and Disease,* ed. W. B. Essman, 2:373–430. New York: Spectrum. 443 pp.
61. Pinnas, J. L., Vanselow, N. A. 1978. Relationship of allergy to headache. *Res. Clin. Stud. Headache* 4:85–95
62. Fernstrom, J. D. 1978. Brain serotonin and nutrition. In *Serotonin in Health and Disease,* ed. W. B. Essman, 3:1–49. New York: Spectrum. 460 pp.
63. Costa, E., Meek, J. 1974. Regulation of the biosynthesis of catecholamines and serotonin in the CNS. *Ann. Rev. Pharmacol.* 14:491–511
64. Aghajanian, G. K. 1972. Influence of drugs on the firing of serotonin-containing neurons in brain. *Fed. Proc.* 31:91–96

Ann. Rev. Med. 1981. 32:73–91
Copyright © 1981 by Annual Reviews Inc. All rights reserved

PAPILLARY AND FOLLICULAR THYROID CANCER: A Selective Approach to Diagnosis and Treatment

❖7383

Ernest L. Mazzaferri, M.D., F.A.C.P.

Department of Medicine, University of Nevada, School of Medicine,
Reno, Nevada 89557

INTRODUCTION

Thyroid carcinoma comprises a group of widely differing tumors ranging from slow-growing, well-differentiated neoplasms that ordinarily have an excellent prognosis to anaplastic cancers that usually have a dismal outcome. This review is concerned with differentiated thyroid cancers—papillary and follicular—the most common types encountered clinically and about which the most controversy exists regarding management.

Benign thyroid nodules are very common, whereas clinically detectable thyroid cancer is relatively rare. Accordingly, a selective approach to the management of thyroid nodules is necessary to avoid subjecting an unacceptably large number of persons to surgery.

When a nodule is found to be papillary or follicular thyroid cancer, opinions differ regarding optimal therapy. Indeed, lively debate continues, not only around the extent of surgery required, but also about the indications for therapeutic ^{131}I and the efficacy of thyroid hormone in controlling tumor growth. This is in part a consequence of the wide range of biologic behavior displayed by these tumors, which makes the task of evaluating therapy difficult (1–3). Papillary and follicular thyroid cancers are often indolent neoplasms with a benign course, including many associated with prior head and neck irradiation (4). Even with distant metastases, patients may live for decades (5, 6). On other occasions death from differentiated thyroid cancer may occur within a few years of diagnosis. An understanding of the clinical behavior and prognosis of these tumors can lead to both an

0066-4219/81/0401-0073$01.00

effective selection of patients for surgery and a reasonable approach to therapy. One can avoid unnecessarily extensive treatment on the one hand and clearly inadequate therapy on the other.

PATHOLOGY AND NATURAL HISTORY

Although disagreement existed in the past concerning the classification of thyroid cancer, in recent years a reasonably uniform system has evolved. The World Health Organization classification (7) is not greatly different from those proposed by the American Thyroid Association (8), the Armed Forces Institute of Pathology (9), or other investigators (10) in which papillary thyroid carcinoma includes both pure papillary and mixed papillary-follicular tumors. Pure follicular carcinoma is classified as a distinct clinicopathologic entity, but both types of cancer arise from the thyroid follicular cell.

Papillary Thyroid Carcinoma

PREVALENCE Papillary thyroid carcinoma accounts for 50–80% of thyroid cancers in the US (1, 3, 10), and inclues almost 90% of all radiation-induced thyroid cancers (4). However, in some European series (11, 12), a substantially larger portion of thyroid cancers are anaplastic.

This neoplasm can appear at any age, but has a peak incidence in the third and fourth decades. It occurs two to three times more frequently in females. In children this ratio is nearly equal, which may in part reflect the influence of radiation-induced disease (13).

PATHOLOGY Papillary thyroid carcinomas vary in size, ranging from microscopic tumors to large cancers that invade the thyroid capsule and grow into contiguous neck structures. Although usually solid, the primary tumor can be hemorrhagic and cystic. These unencapsulated tumors tend to invade lymphatics and spread to the normal surrounding thyroid parenchyma. Multicentricity of the primary tumor is a common feature but varies considerably in different reports, ranging from 20 to 80% depending upon the extent to which the thyroid is examined (1, 10, 14). About half the patients have obvious cervical lymph node metastases at the time of initial surgery (1, 10, 15). A higher frequency of node involvement, ranging up to 75%, is reported on some adult series (16, 17) and in most children with this disease (13, 18). Microscopic metastases may be found in 90% of regional nodes systematically examined (19). Distant metastases to lung, bone, and other soft tissue, in order of decreasing frequency, are seen in fewer than 1% of patients at the time of initial diagnosis (1, 10). However, distant metastases subsequently develop in as few as 4% to as many as 20% (2, 5, 20), with the highest frequency reported in children (18). In a small but significant number of patients, well-differentiated tumors spontaneously undergo transition to anaplastic thyroid carcinoma (21); this appears to be an inherent characteristic of the tumor rather than a consequence of ^{131}I therapy (1).

CLINICAL ASPECTS Papillary thyroid carcinoma, which generally presents as an asymptomatic, slow-growing thyroid nodule, has a prolonged course in most patients. Net mortality from cancer is uncertain, but is perhaps less than 10% over several decades. Cady et al (15) reported mortality to be 10% at 15 years. Woolner (5) reported only 18 deaths (3% mortality) from papillary thyroid carcinoma in 656 patients followed for up to 40 years. We observed only six cancer deaths in 576 patients (1%) followed an average of just over 10 years (2). Even with distant metastases, survival is often prolonged, especially in children and young adults. Harness et al (22) reported 20 of 28 patients (71.4%) with pulmonary metastases to be alive 14 years after ^{131}I therapy. However, Němec et al (6) reported 10- and 15-year survival to be only 29.4 and 14.2%, respectively, in 32 patients with pulmonary metastases. When death does occur from this neoplasm, it may do so as a direct result of local tumor invasion (1).

FACTORS INFLUENCING PROGNOSIS Certain patient features and some tumor characteristics influence prognosis. Children ordinarily have an excellent prognosis (13, 18, 23). A consistent, direct correlation exists between age at the time of diagnosis and mortality from thyroid cancer; the effect of age varies smoothly and gradually from children to those over age 60 (24). Němec et al (6) reported 10-year survival in patients with pulmonary metastases to be 83.3% in those under age 40, whereas none over age 40 survived.

In children, thyroid carcinoma tends to be more aggressive and has a higher mortality rate in males (13). Woolner et al (10) found papillary cancers to be more extensive and aggressive in men but were unable to show differences in survival rates between the sexes. We found recurrence and survival rates to be similar in 576 men and women with this disease; the sizes of the primary lesions, the extent of local invasion, and the presence of regional metastases were also similar in the two groups (1, 2). Others report higher mortality rates in men (15, 24, 25).

A neck mass is the presenting manifestation of papillary thyroid cancer in most patients. We found that if the mass was symptomatic, causing pain, hoarseness, or dysphagia, or if it was growing rapidly, then subsequent tumors recurred more often (21.1 vs 10.7%, $p < .02$) (1, 2). Within limits, the time lapsing from discovery of a thyroid nodule to surgery, averaging 13.0 ± 1.2 (SEM) months, does not influence recurrence or mortality rates (1, 2). Moreover, pregnancy does not alter the course of the disease (1, 26).

The size of the primary lesion has an important impact on prognosis. Small primary tumors have an excellent prognosis. Woolner et al (10) first emphasized the benign clinical behavior of "occult" papillary thyroid carcinoma, defined as a tumor 1.5 cm or less in diameter, with or without regional node metastases. No cancer deaths were observed in 240 patients with occult tumors followed for as long as 40 years (5). In 153 patients with primary tumors of this size, we also observed no cancer deaths over an average of 10 years (1, 2). However, when the primary tumor is larger or invades the thyroid capsule, the disease behaves differently. Woolner et al (5) reported 8 cancer deaths in 348 patients (2.3%) with primary lesions larger than 1.5 cm that did not invade the thyroid capsule. In the same series, when the primary tumor directly invaded the thyroid capsule, cancer deaths occurred in 10 of 68 patients (14.7%). We found when the tumor was large and invaded the thyroid capsule, recurrence was greater (33.3 vs 13.4%, $p < .01$) and death from cancer occurred more often (0.7 vs 7.4%, $p < .03$) than when there was no capsular invasion, although the latter often had regional metastases (1, 2).

Neither the histologic features of this tumor nor the presence of regional lymph node metastases have an important effect upon survival (1, 2, 10, 15). However, more recurrences are seen in those with cervical lymph node metastases, particularly in persons over age 40. (1, and E. L. Mazzaferri, unpublished observation). Although multiple small intrathyroidal tumors are commonly demonstrated, their clinical importance is often questioned. We found almost one third of 70 recurrences to be in the thyroid remnant after no obvious intraglandular dissemination was observed at initial surgery (1, 2).

In summary, the disease is more serious in patients over age 40, but the influence of sex on survival is not consistent. A large primary lesion, the direct invasion of tumor into surrounding neck structures, and the presence of distant metastases are the most important pathologic features imparting a poor prognosis. Bone metastases generally are associated with a shorter survival than pulmonary metastases.

Follicular Thyroid Carcinoma

PREVALENCE Follicular thyroid carcinoma ordinarily accounts for only about 15% of all thyroid cancers (1, 5). It can also be seen at any age, but has a peak occurrence in the third and fourth decades where it is two to three times more frequent in females. Of 214 patients with follicular thyroid cancer, we found the mean age to be 35 ± 12 (SD) years at the time of diagnosis (3). Seventy-two percent of these patients were women.

PATHOLOGY Follicular thyroid cancers are usually encapsulated and have a marked tendancy to vascular invasion. Histologically, these tumors display a spectrum ranging from small, fairly well-differentiated follicles to solid sheets of cells. The tumor most commonly metastasizes to lungs and bone, but can metastasize to brain and other soft tissues, even from a small nonpalpable primary lesion (27). However, this tumor spreads to regional lymph nodes in only 10% of patients (3, 5). Distant metastases occur more frequently from follicular than from papillary thyroid carcinoma. We found 7 of 214 patients (3.3%) with follicular carcinoma to have distant metastases at the time of presentation (3). As compared to people with papillary thyroid carcinoma, more patients with this tumor subsequently develop distant metastases. Němec et al (6) reported that of 87 patients with distant metastases from differentiated thyroid cancer, 32 had papillary and 55 had follicular cancers. Considering the relative frequencies of the two diseases, follicular thyroid cancer is much more likely to metastasize to a distant site than is papillary cancer.

CLINICAL ASPECTS Follicular cancer generally presents as an asymptomatic, slow-growing thyroid nodule. In rare instances, the first clinical manifestations of the disease are due to distant metastases (27). Although net mortality from this disease is uncertain, it is greater than from papillary thyroid cancer. In our series, 5 of 214 patients (2.3%) died of follicular thyroid cancer over a period averaging almost nine years as compared to only six cancer deaths in 576 patients (1.0%) with papillary cancer followed an average of 10 years (2, 3).

FACTORS INFLUENCING PROGNOSIS Increasing age also has an adverse influence on survival with follicular thyroid cancer (6, 15, 24, 25). Distant metastases, especially to bone, portend a poor outcome. Vascular invasion in the primary tumor has an important bearing on prognosis. Woolner et al (5) reported 10- and 20-year survival rates to be 34 and 16%, respectively, in 98 patients with invasive follicular carcinoma as compared to only 3 deaths in 104 patients in whom vascular invasion was judged to be slight or equivocal. We observed a 20% recurrence rate in those with extensive vascular invasion as compared to only 7% when it was slight to moderate (3). The size of the primary tumor and regional lymph node metastases do not have important prognostic implications in patients with follicular cancer (3).

Thus, advancing age of the patient, the presence of vascular and capsular invasion and distant metastases predict a poor outcome with this disease. Bone metastases are associated with a shorter survival than pulmonary metastases (6).

PREVALENCE OF THYROID NODULES AND THYROID CANCER

Thyroid Cancer

Thyroid cancer occurs surprisingly often in carefully examined autopsy specimens, whereas clinically apparent thyroid cancer is relatively rare.

AUTOPSY PREVALENCE Occult thyroid carcinoma is found in up to 28.4% of autopsy specimens from persons exposed to radiation but dying with no known thyroid disease (28). However, Nishiyama et al (29) reported that 13 of 100 consecutive thyroid glands from persons not exposed to radiation and carefully studied at autopsy in Michigan contained small papillary carcinomas. Sampson (28) indicates that these prevalence figures are profound underestimates and suggests that 10–30 million Americans may have occult papillary thyroid cancer.

Most occult carcinomas are much smaller than those originally given this designation by Woolner et al (5). Many are less than 1 mm and most are below 0.5 cm (28). These tumors are not a health problem but must be taken into account when considering therapy.

CLINICAL PREVALENCE Clinically apparent thyroid cancer is not a common disease. In the Third National Cancer Survey in 1969–1971 (30), in the US as a whole for all ages, races, and sexes combined, the age-adjusted incidence for thyroid cancer was 3.7 per 100,000. This represents an increased prevalence over that reported in the Second National Cancer Survey (2.9 per 100,000), which is attributed to the occurrence of radiation-associated neoplasms (4), but it represents only about one new case of thyroid cancer yearly for every 27,000 individuals in the population.

Thyroid Nodules

In contrast to palpable thyroid cancer, thyroid nodules, especially in women, are extremely common. Judging from the Framingham study (31), thyroid nodules, at least half of which are single, occur in about 4% of the adult population. In this long-term study of 5127 persons aged 30–59 years, nontoxic nodules were found in 218 (4.2%) individuals, of whom 183 (6.4%) were women and 35 were (1.5%) men. Maxon et al (32) found that the prevalence of thyroid nodules in the population increased steadily over a wide age range, from about 1% in the second decade of life to about 5%

during the sixth to eighth decades of life. In the Framingham study (31), new nodules, which over 15 years developed at a rate of about 1 per 1000 persons per year, appeared in 0.9% of the men and in 1.7% of the women. None showed evidence of malignancy.

Thyroid nodules are very common, but few become clinically important lesions and even fewer result in death. The clinician must identify and aggressively treat only those that may threaten the patient.

DIAGNOSIS OF THYROID CANCER

Even though most thyroid nodules are benign, at surgery about 20% are generally reported to be cancer, which indicates that a selective process is operative. Routine examination and testing can identify nodules most likely to be malignant.

History and Physical Examination

An asymptomatic neck mass is the presenting feature in 85% of patients with differentiated thyroid cancer (1, 3). Few complain of rapid growth, dysphagia, hoarseness, or pain, symptoms also caused by subacute thyroiditis or anaplastic thyroid cancer.

A question about prior therapeutic head and neck irradiation should be part of any patient's history. External irradiation of the head, neck, or upper thorax, common treatment for a variety of benign conditions from the early 1920s to the late 1950s, is now clearly recognized as being strongly associated with the subsequent development of many thyroid problems (4, 33). Among previously irradiated patients with palpable thyroid abnormalities, the prevalence of thyroid cancer in surgically excised nodules is about 30% (4). Even in an irradiated population, however, benign thyroid nodularity occurs with greater frequency than thyroid cancer. The time required for the development of clinically apparent thyroid cancer following external radiation extends from about 3½ years to an apparent peak 25–35 years (4, 32). The risk of developing thyroid cancer is proportional to the dose of external irradiation and, within the limits of about 10 and 2–3000 rads, is linear (4, 32). Thus, the reason for radiation, the patient's age at the time of irradiation, and the doses and sites of x-ray treatment are important.

One should inquire about family members with thyroid cancer, pheochromocytoma, or hyperparathyroidism. Medullary thyroid carcinoma (MTC), which accounts for about 7% of all thyroid cancers, is a neoplasm of the calcitonin-secreting parafollicular or C cell. This tumor contains amyloid, secretes calcitonin, causes diarrhea, and can be familial when it is transmitted as an autosomal dominant trait. Familial tumors are associated with bilateral pheochromocytomas and hyperparathyroidism (Sipple's syndrome or multiple endocrine neoplasia, type 2). MTC occurs rarely in persons with a phenotype characterized by mucosal neuromas, a typical facial appearance with "bumpy" lips and everted eyelids, a Marfanoid habitus, hypertrophied corneal nerve fibers, and intestinal ganglioneuromas (multiple endocrine neoplasia, type 2b or 3) (2). Once suspected, MTC is excluded by measuring plasma calcitonin after calcium or pentagastrin stimulation (2).

The most important aspect of detecting differentiated thyroid carcinoma is a meticulous examination of the neck. Most patients have a single, small palpable nodule, but a few have a multinodular or diffusely enlarged gland. As judged by palpation, we found the mean diameter of the thyroid nodule to be 2.5 ± 1.4 (SD) cm in patients with papillary thyroid cancer (1). Similarly, 70% of the nodules caused by follicular cancer were less than 3.0 cm in diameter at the time of presentation (3). Cervical lymph nodes are palpable in about one third of those with papillary cancer and in less than 10% of those with follicular cancer (1, 3). One should carefully inspect the neck, especially on swallowing, to note the presence of a mass which might be overlooked by palpation alone. Inspection is facilitated by directing a light across the neck. Palpation of the neck, thyroid gland, cervical lymph nodes, and salivary glands should be carefully done with the patient sitting comfortably. The findings should be recorded by a full-size drawing of the thyroid showing the actual size of the gland and any nodules, along with their relationship to normal anatomic structures in the neck. If there is any question of hoarseness, the vocal cords should be examined by indirect laryngoscopy. Vocal cord paralysis should be considered as a sign of malignancy in any patient with nodular thyroid disease.

Blood Tests

Blood tests are usually unnecessary in the initial examination of a patient with an isolated thyroid nodule unless there is reason to suspect medullary thyroid carcinoma or the patient appears thyrotoxic. In patients with a past history of head and neck irradiation, the frequency of positive tests for antithyroglobulin antibodies, antimicrosomal antibodies, and serum thyroglobulin is significantly increased above that found in a control population; however, these tests will not identify those with thyroid cancer (34). Patients with an autonomously hyperfunctioning thyroid nodule are usually euthyroid but may be thyrotoxic, and they ordinarily show nonsuppressable radionuclide uptake in the nodule following thyroid hormone feeding (35).

Thyroid Scans

Thyroid imaging is indicated in the patient with an isolated thyroid nodule and should be done with either ^{99m}Tc pertechnetate or ^{123}I. Both deliver much less radiation to the thyroid than ^{131}I (36). A gamma counter with a pinhole collimator is the most sensitive imaging device. Most thyroid cancers fail to concentrate radionuclide to any extent and appear as hypofunctional ("cold") nodules, whereas most hyperfunctional ("hot") nodules are benign. There are, however, limitations to these generalizations that must be taken into account when making therapeutic decisions.

Some prefer ^{99m}Tc as the scanning agent since it delivers the lowest radiation dose to the thyroid. However, ^{123}I may provide additional information about the functional capacity of the nodule because it is trapped, organified, and incorporated into thyroid hormone; ^{99m}Tc is merely trapped but is not organified. So ^{99m}Tc scans may sometimes show a nodule to be hyperfunctional which by radioiodine scan is hypofunctional. The opposite may also be true (37). Although hyperfunctioning nodules are almost invariably benign, there are well-documented exceptions (38). Thyroid cancers, while most commonly hypofunctional, may give the appearance of concentrating radionuclide, presumably because they are small or surrounded by normal thyroid tissue. Moreover, most hypofunctional nodules are benign.

Thyroid scanning alone is of relatively little discriminatory value in the identification of thyroid cancer (1). Nelson et al (39) correlated thyroid scans with the surgically documented size, location, and histology of thyroid cancers in 67 patients and concluded that, unless the nodule is hyperfunctioning, clinical criteria rather than the scan should contribute most in deciding whether to treat surgically.

Disagreement exists regarding thyroid scanning of previously irradiated patients in whom the thyroid is palpably normal. Favus et al (40) palpated thyroid lesions in only 16.5% of 1056 irradiated persons, while an additional 10.7% had defects detected by ^{99m}Tc scanning. Of the 60 cancers resected in this group, one third were smaller than 0.6 cm and were not detected preoperatively either by scan or by palpation, but were incidental small cancers found while resecting a benign nodule. Over two thirds of the 60 cancers were 1.5 cm or smaller in diameter, lesions defined as occult by Woolner et al (5), from which there is no apparent mortality. A subsequent report (41) indicated that of 68 patients with palpably normal thyroids or equivocal physical findings, 18.2% of the former and 19.6% of the latter had malignancies detected by abnormal scan alone. A gamma camera with a pinhole collimator can detect thyroid lesions as small as 0.5–0.8 cm, whereas most experienced examiners can not palpate thyroid nodules much smaller than 1.0 cm. The central issue is, however, whether one must detect lesions smaller than 1.0 cm since most are papillary cancers with, at this size, a benign course.

Thyroid Ultrasonography

The palpable, isotopically hypofunctional or normally functioning nodule should be further evaluated by echography, which ordinarily can resolve a lesion of about 2 cm in diameter and will show it to be a solid, fluid-filled, or mixed lesion. Thyroid cysts appear as globular, smooth-walled structures without internal echos; they account for about 15–20% of all isolated thyroid nodules and are rarely malignant (42). Only about 1–2% of primary thyroid carcinomas appear cystic, and are usually mixed nodules representing degenerative tumors. Metastatic lymph nodes from differentiated thyroid cancer are more often cystic lesions. If the lesion is a simple cyst, it should be aspirated. Clark et al (43) reported that, among 35 patients with cystic lesions treated by aspiration, 40% had a sustained remission, 29% had a substantial regression and 31% had unsuccessful aspiration with recurrence of fluid. Operation was unnecessary in about 70% of these patients. Solid and mixed lesions must be regarded as being potentially malignant. Thyroid echography is also of value in following changes in the size and configuration of nodules during thyroid hormone treatment. Used in conjunction with needle aspiration, this is a particularly safe and highly reliable procedure.

Thyroid Needle Biopsy

During the past decade in which thyroid needle biopsy has been used fairly routinely in the assessment of thyroid nodules in many centers, there have been no reports of implantation of cancer in the needle tract or other serious complications (44). However, regardless of the type of biopsy taken (core, cell block, or cytology), papillary and anaplastic cancers are ordinarily easily recognized, but follicular carcinoma is often difficult to distinguish from a benign adenoma.

Löwhagen et al (45) in Sweden used fine-needle aspiration biopsy for cytology to examine 412 patients, all of whom were subsequently operated. There were no false-positive cytologic diagnoses and there was a high degree of accuracy in recognizing or suspecting thyroid cancer. Walfish et al (46) in Canada performed fine-needle (25–27 gauge) aspiration biopsy in 82 lesions that were predominantly solid by echography. Although they were usually able to accurately predict the lesion found at surgery, in 4 of 65 nodules (6%) excised, the cytology was read as benign but the nodule was shown to be malignant at surgery. Crile et al (44) performed biopsy in 576 thyroids from 1971 through 1978, and had only one case of cancer appear a year after biopsy, which was an anaplastic cancer arising from a different part of the gland than had been biopsied. Hamburger et al (47) reported that of 116 surgically confirmed thyroid malignancies, only 62 (59%) were recognized as such on the basis of clinical findings; but on the basis of fine-needle aspiration, 107 (92%) were considered malignant preoperatively and another 8 (7%) were suspected of being malignant. However, 7 malignancies and 2 atypical adenomas were thought to be benign by aspiration cytology preoperatively. Since it is difficult to distinguish adenomas from follicular cancers, it may be necessary to remove all firm nodules that are highly cellular by biopsy. At the Cleveland Clinic in 1978, of all patients with palpable thyroid lesions, 132 (82%) had cancer excluded by needle biopsy and were treated medically instead of surgically (44). Of 910 patients examined by Hamburger et al (47), the proportion of patients for whom surgery was advised was reduced, on the basis of aspiration biopsy, from 65 to 30%. The most limiting aspect of this technique is the availability of an experienced pathologist trained in the interpretation of thyroid cytology.

Thyroid Hormone

Thyroid nodules often decrease in size or disappear after TSH suppression. DePapp et al (48), treating 35 previously irradiated persons with thyroxine, found that 9 single nodules decreased in size and 2 disappeared completely. Although, on occasion, malignant nodules may initially shrink (1, 48), most either grow or remain unchanged during thyroid hormone suppression of TSH. Shimaoka (49) reported that, overall, about 64% of 114 patients with thyroid nodules experienced a reduction in nodule size with thyroid hormone therapy, and that liothyronine (T_3) was more effective than L-thyroxine (T_4) in this regard. Thyroid nodules in about one third of the responders became either nonpalpable or unmeasurable. Surgery was done in 11 of 45 (24%) irradiated individuals and in 11 of 62 (18%) nonirradiated persons with thyroid nodules that did not regress. Three cancers were found in each group.

Summary of Diagnostic Recommendations

The order in which tests are done depends in part on the examiner. After the initial evaluation, one usually obtains a thyroid radionuclide scan. If the nodule is not hyperfunctional, thyroid echography is done. Simple cysts are aspirated. Mixed and solid lesions should be further evaluated, either with aspiration biopsy or with thyroid hormone suppression. If the nodule is malignant or suspicious by aspiration biopsy, or if it fails to completely regress or grows after about 6–12 months of thyroid hormone therapy, it should be excised. Older persons, particularly women, can usually be safely treated with thyroid hormone for long periods if the nodule remains stable. If possible, aspiration biopsy should be done before long-term suppression therapy is instituted. Ordinarily, small thyroid nodules ($<$ 1.5 cm) associated with prior head and neck irradiation can be similarly managed, whereas large solid nodules usually should be promptly excised.

THERAPY OF PAPILLARY AND FOLLICULAR THYROID CARCINOMA

Initial treatment generally consists of surgery, often followed by ^{131}I therapy, and invariably includes thyroid hormone.

Surgery

Some surgeons consider lobectomy with regional lymph node dissection adequate initial management (50, 51). However, Shands & Gatling (52) reported an 11% contralateral recurrence rate after subtotal thyroidectomy; and half of these patients died. Tollefsen et al (53) reported contralateral thyroid recurrence to be only 4.6%, but almost half died from recurrent disease.

We observed 84 recurrences in 571 patients with papillary thyroid carcinoma following initial therapy. Recurrences following subtotal thyroidectomy were almost twice as frequent as those following total thyroidectomy (19.2 vs 11.0%, $p < .01$) (2). With smaller primary lesions ($\leqslant$ 1.5 cm) fewer recurrences were observed after total thyroidectomy than after subtotal thyroidectomy (2.8 vs 6.2%), but the difference was not statistically significant (E. L. Mazzaferri, unpublished information). Sites of recurrence were cervical lymph nodes alone (58), remaining thyroid alone (10), lymph nodes and thyroid (13), lung (6), lung and bone (2), brain (1). Perhaps of most importance, five of six cancer deaths occurred in patients with recurrent disease.

Of 182 follicular carcinoma patients in our study considered free of disease after initial therapy, 20 experienced recurrences; 8 in the opposite lobe, 10 in regional lymph nodes, and 2 were distant recurrences to lung and bone (3). Although the extent of thyroid surgery did not substantially alter the recurrence rate, almost half occurred in the remaining thyroid.

In patients previously exposed to radiation, Block (54) observed about a 10% recurrence in the remaining thyroid. He advises total thyroidectomy in most patients with large papillary carcinomas and invasive follicular cancers but not for isolated, small papillary cancers or low-grade follicular cancers. Most agree with this approach for radiation-induced thyroid cancer (55).

Few advocate radical neck dissection, but some perform modified neck dissection for node metastases from differentiated thyroid cancer. We found recurrences following no lymph node dissection, simple excision of involved nodes, and neck dissection to be 12.3, 17.9, and 14.7% (p = n.s.) respectively (2).

Radioactive Iodine

Following treatment with ^{131}I, Varma et al (56) reported improved survival only in patients over age 40, whereas Leeper (57) showed enhanced survival only in patients younger than 40 years. Krishnamurthy & Blahd (58), in a 25-year prospective study of 54 patients with thyroid cancer of various histologic types, found the response to ^{131}I ablation to be the most critical factor influencing survival. None of 44 patients in whom total ^{131}I ablation of primary and metastatic foci was achieved died of cancer. Of 10 others in whom total ablation was not possible, 7 died of thyroid carcinoma and, in 5 of these patients, it was a differentiated cancer.

In our study, 116 patients with papillary cancer were initially given ^{131}I to treat cervical metastases (80 patients), to treat pulmonary metastases (3 patients), or to ablate remaining radioiodine uptake in the thyroid bed (33 patients) (1). Recurrences were significantly fewer ($p < .001$) in those treated with ^{131}I (Figure 1). In the most recent analysis, the statistical difference in recurrence between ^{131}I and thyroid hormone alone persisted only for patients with primary tumors larger than 1.5 cm (E. L. Mazzaferri, unpublished data). Of 214 patients with follicular thyroid carcinoma, the lowest recurrence rate (4.7%) was observed in 42 patients treated with total thyroidectomy, ^{131}I and thyroid hormone (3).

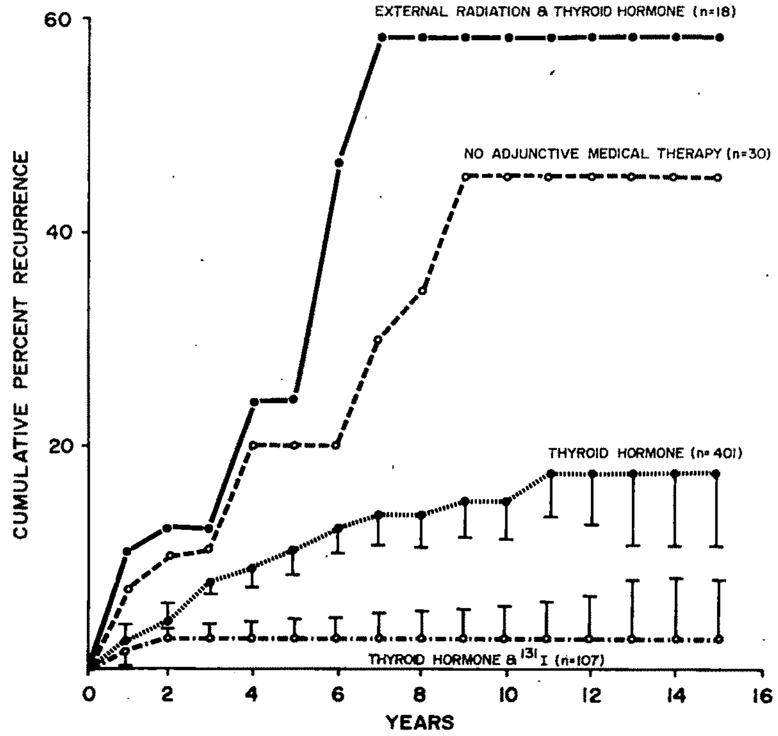


Figure 1 Cumulative recurrence rate following various forms of adjunctive therapy. From Mazzaferri et al (1), with permission.

In 66 patients with distant metastases, Němec et al (6) found ^{131}I uptake in 56% of papillary and 46.3% of follicular metastases. Uptake of ^{131}I was achieved more frequently in younger patients, in those with papillary thyroid cancer, and in persons with fine, "snow-flake" pulmonary metastases on chest x-ray. Tumor uptake of ^{131}I is critical since 10-year survival for papillary carcinoma was 83.3% and zero, and for follicular carcinoma was 64.3% and 8.3%, depending upon whether the tumor did or did not concentrate ^{131}I.

The belief that thyroid hormone alone is sufficient for the treatment of pulmonary metastases (59) is unique and not shared by most (1, 6, 56–58). The danger of leukemia (60), the possibility of causing pulmonary fibrosis, and the risk of producing anaplastic thyroid cancer or genetic damage are the usual reasons given for withholding ^{131}I and treating with thyroid hormone alone (59). However, with current recommendations regarding the timing and doses of ^{131}I (61), these appear not to be substantial risks in most patients. First, treatment does not produce pulmonary fibrosis unless the patient accumulates 100 mCi or more in his lungs, which can rarely, if ever, be achieved with the maximum allowable single dose of 200 mCi (61). Second, leukemia, reported after larger doses of ^{131}I often given at several month intervals (60), is not reported with increased frequency in patients receiving lower doses given at about yearly intervals (1, 3, 56–58, 61). Third, anaplastic carcinoma occurs in the absence of irradiation as part of the natural history of differentiated thyroid carcinoma (1, 21). Fourth, long-term follow-up of children treated for thyroid cancer with ^{131}I discloses no apparent increase in the occurrence of infertility or birth defects (62).

Thyroid Hormone

There is abundant evidence that TSH can stimulate differentiated thyroid cancers to grow. Recurrence rates of both papillary and follicular thyroid cancer are lower in patients treated postoperatively with thyroid hormone (Figure 1) (1–3). Thyroid hormone given in adequate doses suppresses TSH to low or undetectable levels that will not rise in response to TRH (63).

Summary of Therapeutic Recommendations

Minimum therapy for small isolated nonmetastatic papillary cancer is a total lobectomy and isthmusectomy on the affected side with a subtotal lobectomy on the contralateral side. Papillary cancer that is larger ($>$ 1.5 cm), bilateral, metastatic, or invades the thyroid capsule, and most follicular cancers should be treated with total or near-total thyroidectomy followed by total ablation of all remaining tissue that concentrates ^{131}I. Lymph nodes should be excised, but extensive neck dissection is rarely, if ever, indicated.

A total body ^{131}I scan is done with 2–5 mCi six weeks postoperatively. Since normal thyroid tissue traps ^{131}I with perhaps a 20-fold greater avidity than even the most differentiated forms of thyroid cancer, ^{131}I scans are inaccurate if substantial amounts of normal thyroid tissue are left in place. High TSH levels and a relatively low total body iodine pool maximize tumor uptake of ^{131}I. Intramuscular bovine TSH can be given prior to the scan, but this results in added expense, side effects that are occasionally severe (64), and fluctuating serum TSH levels (65). Endogenous TSH is best raised by giving liothyronine (T_3) three times daily for four weeks then discontinuing it for two weeks before the scan (66). Using this method, Martin (67) found TSH levels after thyroidectomy to be about 20 μU/ml in all but one cancer patient and found no qualitative differences in scans performed after bovine TSH injections or T_3 withdrawal. Edwards et al (68) studied 36 patients after thyroid hormone withdrawal and concluded that, in general, a tumor can not be considered incapable of concentrating ^{131}I until serum TSH levels exceed 30μU/ml. A low-iodine diet should also be given prior to the scan. Goslings (69) demonstrated the mean ^{131}I tumor dose to double as a result of a low-iodine diet given for five days. Lymph node and lung metastases can be ablated in many patients given a single 100-mCi treatment dose of ^{131}I after the use of a low-iodine diet (61). Certain diuretics will deplete iodine stores but are usually not necessary.

If substantial ^{131}I uptake is demonstrable in the thyroid bed, it is ablated with ^{131}I. Although lower doses (<30 mCi) can be used for this purpose (70), about one third or more patients must be given 100–150 mCi ^{131}I (61), which requires hospitalization for 3–5 days. Patients with significant retention of ^{131}I in metastatic deposits are routinely treated with 150 mCi ^{131}I, which can be given without inducing radiation sickness and will ordinarily not produce damage to any structure, including lungs. Bone metastases are usually treated with 200 mCi. Treatment should be continued with ^{131}I until all residual uptake is ablated or adverse effects of ^{131}I are detected. Individual doses are usually given at about 6–12-month intervals and should not be administered before the bone marrow has fully recovered from the previous dose. Leukemia occurs with increased frequency in patients who receive multiple, large doses of ^{131}I given at frequent intervals (60).

Following initial therapy, if the patient is shown by scan to be free of disease, scans can usually be repeated at 3- and then 5-year intervals. Although most differentiated thyroid cancers recur within the first decade after initial treatment, late recurrences during the second decade can be seen (1–3, 60).

Serum thyroglobulin (TG) measurements should be made after surgery and at yearly intervals thereafter. Van Herle & Uller (71) found serum TG levels in untreated patients with differentiated thyroid cancer to range from 22 to 445 ng/ml (average 144.3 ± 46.5) as compared to a range of 0–20.7

ng/ml in 95 normal subjects. TG levels returned to normal in successfully treated patients, but remained elevated in others with metastases. LoGerfo et al (72) found serum TG levels to be less than 10 ng/ml in disease-free patients as compared to levels above 100 ng/ml in patients with metastatic disease. Elevated serum TG levels, which are not a specific marker of cancer, can also be seen in patients with Graves' disease or thyroiditis and following prior head and neck irradiation in the absence of detectable tumor. However, they serve as a simple and effective means of following the patient indefinitely.

CONCLUSION

Using a selective approach, surgeons can excise thyroid nodules most likely to be malignant. Minimum therapy of small isolated nonmetastatic papillary cancer is lobectomy, isthmusectomy, and contralateral subtotal lobectomy. Papillary cancers that are large (> 1.5 cm), bilateral, metastatic, or locally invasive and most follicular cancers should be treated with total or near-total thyroidectomy followed by total ablation of remaining tissue that concentrates ^{131}I. In most instances, lymph nodes should be simply excised. Thyroid hormone suppression of TSH should always be done. Survival is usually excellent except when the tumor is locally invasive or is metastatic to a distant site. Mortality is substantially reduced if the tumor concentrates ^{131}I and residual tumor ^{131}I uptake can be completely ablated.

Literature Cited

1. Mazzaferri, E. L., Young, R. L., Oertel, J. E., Kemmerer, W. T., Page, C. P. 1977. Papillary thyroid carcinoma: the impact of therapy in 576 patients. *Medicine* 56:171–96
2. Mazzaferri, E. L. 1978. Thyroid cancer: an overview. *Thyroid Today* 1(10):1–6
3. Young, R. L., Mazzaferri, E. L., Rahe, A. J., Dorfman, S. G. 1980. Pure follicular thyroid carcinoma: impact of therapy in 214 patients. *J. Nucl. Med.* 21:733–37
4. Mazzaferri, E. L. 1980. Thyroid carcinoma following therapeutic and accidental radiation exposure. In *Special Topics in Endocrinology and Metabolism,* ed. M. Cohen, P. P. Foa. New York: Liss. In press
5. Woolner, L. 1971. Thyroid carcinoma: pathologic classification with data on prognosis. *Semin. Nucl. Med.* 1:481–502
6. Němec, J. V., Zamrazil, V., Pohunková, D., Röhling, S. 1979. Radioiodine treatment of pulmonary metastases of differentiated thyroid cancer: results and prognostic factors. *Nucl. Med.* 18:86–90
7. Histologic typing of thyroid tumors. 1974. In *International Histological Classification of Tumours,* No. 11, p. 22. Geneva: World Health Organ. 28 p.
8. Beierwaltes, W., DeGroot, L. J., Feind, C. R., Kriss, J. P., Meissner, W. A., Werner, S. C. 1969. Classification of thyroid disease. Report of the Committee on Nomenclature of the American Thyroid Association. *J. Clin. Endocrinol. Metab.* 29:860–62
9. Meissner, W. A., Warren, S. 1969. Tumors of the thyroid gland. In *Atlas of Tumor Pathology,* pp. 55–59, Second Series, Fascicle 4. Washington, DC: Armed Forces Inst. Pathol. 135 p.
10. Woolner, L. B., Beahrs, O. H., Black, W. M., McConahey, W. M., Keating, F. R. Jr. 1961. Classification and prognosis of thyroid carcinoma. A study of 885 cases observed in a thirty year period. *Am. J. Surg.* 102:354–89

11. Wade, J. S. H. 1969. Review of thyroid carcinoma in the United Cardiff Hospitals. *Proc. R. Soc. Med.* 62:1–3
12. Lindahl, F. 1976. Anaplastic thyroid carcinoma in Denmark 1943–1968. *Dan. Med. Bull.* 23:120–28
13. Winship, T., Rosvoll, R. V. 1970. Thyroid carcinoma in childhood: final report on a 20-year study. *Clin. Proc. Child. Hosp. Natl. Med. Cent.* 26:327–48
14. Russell, W. O., Ibanez, M. L., Clark, R. L., White, E. C. 1963. Thyroid carcinoma: classification, intraglandular dissemination, and clinicopathologic study based upon whole organ sections of 80 glands. *Cancer* 16:1425–60
15. Cady, B., Sedgwick, C. E., Meissner, W. A., Bookwalter, J. R., Romagosa, V., Weber, J. 1976. Changing clinical, pathologic, therapeutic and survival patterns in differentiated thyroid carcinoma. *Ann. Surg.* 183:541–53
16. Crile, G. Jr. 1971. Changing end results in patients with papillary carcinoma of the thyroid. *Surg. Gynecol. Obstet.* 132:460–65
17. Mustard, R. A. 1975. The role of surgery in the management of thyroid cancer. *Can. Med. Assoc. J.* 113:109–13
18. Harness, J. K., Thompson, N. W., Nishiyama, R. H. 1971. Childhood thyroid carcinoma. *Arch. Surg.* 102:278–84
19. Noguchi, S., Noguchi, A., Murakami, N. 1970. Papillary carcinoma of the thyroid. II. Value of prophylactic lymph node excision. *Cancer* 26:1061–64
20. Noguchi, S., Noguchi, A., Murakami, N. 1970. Papillary carcinoma of the thyroid. I. Developing pattern of metastases. *Cancer* 26:1053–60
21. Nishiyama, R. H., Dunn, E. L., Thompson, N. W. 1972. Anaplastic spindle-cell and giant-cell tumors of the thyroid gland. *Cancer* 30:112–27
22. Harness, J. K., Thompson, N. W., Sisson, J. C., Beierwaltes, W. H. 1974. Differentiated thyroid carcinoma. Treatment of distant metastases. *Arch. Surg.* 108:410–19
23. Liechty, R. D., Safaie-Shirazi, S., Soper, R. T. 1972. Carcinoma of the thyroid in children. *Surg. Gynecol. Obstet.* 134:595–99
24. Halnan, K. E. 1966. Influence of age and sex on incidence and prognosis of thyroid cancer: 344 cases followed for ten years. *Cancer* 19:1534–41
25. Cady, B., Sedgwick, C. E., Meissner, W. A., Wool, M. S., Ferdinand, S. A., Weber, J. 1979. Risk factor analysis in differentiated thyroid cancer. *Cancer* 43:810–20
26. Rosvoll, R. V., Winship, T. 1965. Thyroid carcinoma and pregnancy. *Surg. Gynecol. Obstet.* 121:1039–42
27. Boehm, T., Rothouse, L., Wartofsky, L. 1976. Metastatic occult follicular thyroid carcinoma. *J. Am. Med. Assoc.* 235:2420–21
28. Sampson, R. J. 1977. Prevalence and significance of occult thyroid cancer. In *Radiation-Associated Thyroid Carcinoma,* ed. L. J. DeGroot, L. A. Frohman, E. L. Kaplan, S. Refetoff, pp. 137–53. New York, San Francisco, London: Grune & Stratton. 539 pp.
29. Nishiyama, R. H., Ludwig, G. K., Thompson, N. W. 1977. See Ref. 28, pp. 123–35
30. Biometry Branch, Division of Cancer Cause and Prevention. 1975. In *Third National Cancer Survey. Incidence Data.* ed. S. J. Cutler, J. L. Young, p. 25. Washington, DC: DHEW Publ. No. (NIH) 75-787, NCI Monogr. No. 41, March. 454 p.
31. Vander, J. B., Gaston, E. A., Dawber, T. R. 1968. The significance of nontoxic thyroid nodules. *Ann. Intern. Med.* 69:537–40
32. Maxon, H. R., Thomas, S. R., Saenger, E. L., Buncher, C. R., Kereiakes, J. G. 1977. Ionizing irradiation and the induction of clinically significant disease in the human thyroid gland. *Am. J. Med.* 63:967–78
33. Hempelmann, L. H., Hall, W. L., Phillips, M., Cooper, R. A., Ames, W. R. 1975. Neoplasms in persons treated with x-ray in infancy: 4th survey in 20 years. *J. Natl. Cancer Inst.* 55:519–80
34. DeGroot, L. J., Hoge, K., Refetoff, S., Van Herle, A. J., Asteris, G. T., Rochman, H. 1977. Serum antigens and antibodies in the diagnosis of thyroid cancer. *J. Clin. Endocrinol. Metab.* 45:1220–23
35. Hamburger, J. I. 1975. Solitary autonomously functioning thyroid lesions. Diagnosis, clinical features and pathogenetic considerations. *Am. J. Med.* 58:740–48
36. Silverman, C., Hoffman, D. A. 1975. Thyroid tumor risk from radiation during childhood. *Prev. Med.* 4:100–5
37. Barth, J. D., Bakker, W. H., Henneman, G. 1978. Discrepancies between iodine and technetium thyroid scintigraphy. *J. Am. Med. Assoc.* 240:463–64
38. Lamberg, B-A., Makinen, J., Murtoma, M. 1976. Papillary thyroid carcinoma

in toxic adenoma. *Nucl. Med.* 15:138–41

39. Nelson, R. L., Wahner, H. W., Gorman, C. A. 1978. Rectilinear thyroid scanning as a predictor of malignancy. *Ann. Intern. Med.* 88:41–44
40. Favus, M. J., Schneider, A. B., Stachura, M. E., Arnold, J. E., Ryo, U. Y., Pinsky, S. M., Colman, M., Arnold, M. J., Frohman, L. A. 1976. Thyroid cancer occurring as a late consequence of head-and-neck irradiation. Evaluation of 1056 patients. *N. Engl. J. Med.* 294:1019–25
41. Pinsky, S. M., Ryo, U. Y. 1977. Technique and utility of thyroid scans. See Ref. 28, pp. 297–309
42. Clark, O. H., Greenspan, F. S., Coggs, G. C., Goldman, L. 1975. Evaluation of solitary cold nodules by echography and thermography. *Am. J. Surg.* 130:206–10
43. Clark, O. H., Okerlund, M. D., Cavalieri, R. R., Greenspan, F. S. 1979. Diagnosis and treatment of thyroid, parathyroid, and thyroglossal duct cysts. *J. Clin. Endocrinol. Metab.* 48:983–88
44. Crile, G. Jr., Esselstyn, C. B. Jr., Hawk, W. A. 1979. Needle biopsy in the diagnosis of thyroid nodules appearing after radiation. *N. Engl. J. Med.* 301:997–98
45. Löwhagen, T., Granberg, P.-O., Lundall, G., Skinnari, P., Sunblad, R., Willems, J.-S. 1979. Aspiration biopsy cytology (ABC) in nodules of the thyroid gland suspected to be malignant. *Surg. Clin. North Am.* 59:3–8
46. Walfish, P. E., Hazarri, E., Strawbridge, H. T. G., Miokin, M., Rosen, I. B. 1977. An evaluation of combined ultrasound and needle aspiration cytology in the assessment and management of the hypofunctioning thyroid nodule. *Ann. Intern. Med.* 87:270–74
47. Hamburger, J. I., Miller, J. M., Kini, S. R. 1979. *Clinical-Pathological Evaluation of Thyroid Nodules. Handbook and Atlas.* pp. 75–80. Southfield, Mich: Private publication. 93 p.
48. DePapp, Z. G., Pincus, R. A., Hempelmann, L. H. 1970. Treatment of radiation-induced nodular goiter. *J. Nucl. Med.* 11:496–501
49. Shimaoka, K. 1977. Suppressive therapy with thyroid hormones. See Ref. 28, pp. 433–44
50. Buckwalter, J. A., Thomas, C. G., Freeman, J. B. 1975. Is childhood cancer a lethal disease? *Ann. Surg.* 181:632–37
51. Tiel, A. 1970. Carcinoma of the thyroid. *Proc. R. Soc. Med.* 58:309–10
52. Shands, W. B., Gatling, R. R. 1970. Cancer of the thyroid: review of 109 cases. *Ann. Surg.* 171:735–41
53. Tollefsen, H. R., Shah, J. P., Huvos, A. G. 1972. Papillary carcinoma of the thyroid: recurrence in the gland after initial surgery treatment. *Am. J. Surg.* 124:468–74
54. Block, M. A. 1977. Surgery of the irradiated thyroid gland for possible carcinoma: criteria, technique, results. See Ref. 28, pp. 353–70
55. Kaplan, E. L. 1977. Operation of the irradiated gland for possible thyroid carcinoma: criteria, technique and results. See Ref. 28, pp. 371–82
56. Varma, V. M., Beirwaltes, W. H., Nofal, M., Nishiyama, R. H., Copp, J. R. Jr. 1970. Treatment of thyroid cancer: death rates after surgery and after surgery followed by sodium iodine I^{131}. *J. Am. Med. Assoc.* 214:1437–42
57. Leeper, R. D. 1973. The effect of ^{131}I therapy on survival of patients with metastatic papillary or follicular thyroid carcinoma. *J. Clin. Endocrinol. Metab.* 36:1143–49
58. Krishnamurthy, G. T., Blahd, W. H. 1977. Radioiodine I-131 therapy in the management of thyroid cancer. A prospective study. *Cancer* 40:195–202
59. Crile, G. Jr. 1966. Endocrine dependency of papillary carcinoma of the thyroid. *J. Am. Med. Assoc.* 195:721–24
60. Brickner, H., Hansen, H. S., Anderson, A. P. 1973. Induction of leukemia by ^{131} I treatment of thyroid carcinoma. *Br. J. Cancer* 28:232–38
61. Beierwaltes, W. H. 1978. The treatment of thyroid carcinoma with radioactive iodine. *Semin. Nucl. Med.* 8:79–93
62. Sarkar, S. D., Beierwaltes, W. H., Gill, S. P., Cowley, B. J. 1976. Subsequent fertility and birth histories in children and adolescents treated with ^{131}I for thyroid cancer. *J. Nucl. Med.* 17:460–64
63. Busnardo, B., Vangelista R., Girelli, M. E., Bui, F., Lazzi, C. 1976. TSH levels and TSH response to TRH as a guide to the replacement treatment of patients with thyroid carcinoma. *J. Clin. Endocrinol. Metab.* 42:901–6
64. Krishnamurthy, G. T. 1978. Human reaction to bovine TSH: concise communication. *J. Nucl. Med.* 19:284–86
65. Hershman, J. M., Edwards, C. L. 1972. Serum thyrotropin (TSH) levels after thyroid ablation compared with TSH levels after exogenous bovine TSH: implications for ^{131}I treatment of thyroid

carcinoma. *J. Clin. Endocrinol. Metab.* 34:814–18

66. Stein, D. B., Nicoloff, J. T. 1971. Triiodothyronine withdrawal test—a test of thyroid-pituitary adequacy. *J. Clin. Endocrinol. Metab.* 32:127–29
67. Martin, N. D. 1978. Endogenous serum TSH levels and metastatic survey scans in thyroid cancer patients using triiodothyronine withdrawal. *Clin. Nucl. Med.* 10:401–3
68. Edwards, C. J., Hayes, S., Kermode, J. C., Thompson, B. D. 1977. Measurement of serum TSH and thyroid hormone in the management and treatment of thyroid carcinoma with radioiodine. *Br. J. Radiol.* 50:799–802
69. Goslings, B. M. 1975. Effect of a low iodine diet on ^{131}I therapy in follicular thyroid carcinomata. Proceedings of the Society of Endocrinology. *J. Endocrinol.* 30P:40 (Abstr.)
70. McGowen, K. D., Adler, R. A., Ghaed, N., Verdon, T., Hofeldt, F. D. 1976. Low dose radioiodide ablation in postsurgical patients with thyroid cancer. *Am. J. Med.* 61:52–58
71. Van Herle, A., Uller, R. P. 1975. Elevated serum thyroglobulin. A marker of metastases in differentiated thyroid carcinomas. *J. Clin. Invest.* 56:272–77
72. LoGerfo, P., Stillman, T., Colacchio, D., Feind, C. 1977. Serum thyroglobulin and recurrent thyroid cancer. *Lancet* 1:881–82

Ann. Rev. Med. 1981. 32:93–99
Copyright © 1981 by Annual Reviews Inc. All rights reserved

GENETIC COMPONENT OF ALCOHOLISM

❖7384

Donald W. Goodwin, M.D.

Department of Psychiatry, University of Kansas Medical Center,
39th and Rainbow Boulevard, Kansas City, Kansas 66103

Family studies of alcoholism consistently emphasize its high prevalence among the relatives of alcoholics (1). In most studies, at least 25% of male relatives are alcoholic. As a rule, these studies also show a higher-than-expected frequency of alcoholism among female relatives (between 5 and 10%). Population surveys indicate that between 3 and 5% of male adults and between 0.1 and 1% of female adults are alcoholic (2).

"Familial" of course is not synonymous with genetic. Speaking Chinese runs in families, but not because of genes. In humans, there are three strategies for identifying hereditary factors in family illnesses: (*a*) twin studies, (*b*) genetic marker studies, and (*c*) adoption studies. All have been applied to alcoholism.

Twin Studies

There have been four twin studies of alcoholism. One, in Sweden, found that identical twins were concordant for alcoholism more often than were fraternal twins (3). The more severe the alcoholism, the greater the discrepancy between concordance rates in identical vs fraternal twins.

Another study, in Finland, found that identical twins were more concordant for quantity and frequency of drinking but not for adverse consequences of drinking (4). There was a trend for younger identical twins to be more concordant for adverse consequences (which may not be coincidental, see below).

In 1972, Loehlin analyzed questionnaire data from 850 like-sex twins in the United States (unpublished data). His findings were consistent with a genetic factor in drinking behavior. Jonsson & Nilsson (5) reported findings

0066-4219/81/0401-0093$01.00

also based on questionnaire data from 7500 twin pairs in Sweden. Monozygotic twins were more concordant with regard to quantity of alcohol consumed, but not with regard to adverse consequences. The results resemble those of the Finnish study (4).

All these twin studies point to genetic control of drinking behavior, but the evidence for a genetic determinant of alcoholism is inconsistent. Other twin studies (6) indicate genetic control over the metabolism of alcohol, but varying rates of metabolism probably have no relevance to alcoholism (7).

Genetic Marker Studies

Establishment of a relationship between a known inherited biological trait and a familial disease would also suggest that the latter was genetic. More than a score of genetic marker studies relevant to alcoholism have been published (4). Almost without exception, whenever a report of an association between a marker and alcoholism is followed by attempts to replicate the finding, the findings are contradictory.

Among markers studied have been color blindness, blood groups, genetically determined proteins, and finger ridge count. The observation (8) that alcoholics are likely to be color blind has been confirmed in several studies (9), but apparently the color blindness is reversible, perhaps caused by malnutrition or toxic effects.

The Danish Adoption Studies

A third approach to separating "nature" from "nurture" is to study individuals separated from their biological relatives soon after birth and raised by nonrelative foster parents.

Beginning in 1970, my colleagues and I started a series of adoption studies in Denmark, supported by the National Institute of Alcohol Abuse and Alcoholism, and intended to investigate further the possibility that alcoholism in part has genetic roots. The studies have been completed and the results published (10–13). This section summarizes these results and the next section discusses in more detail other adoption studies.

The Danish adoption studies involved interviewing four groups of subjects. The first group consisted of sons of alcoholics (average age, 30 years) raised by nonalcoholic foster parents. The second group consisted of sons of alcoholics (average age, 33 years) raised by their alcoholic biological parents. The third and fourth groups consisted, respectively, of daughters of alcoholics (average age, 37 years) raised by nonalcoholic foster parents and daughters of alcoholics (average age, 32 years) raised by their alcoholic biological parents. Paired with each group was a control group matched for age and, in the adopted samples, circumstances of adoption. All adoptees were separated from their biological parents in the first few weeks of life and adopted by nonrelatives. The interviews were conducted by Danish psychiatrists "blind" to the overall purpose of the study and the identity of the interviewees, that is, whether they were children of alcoholics or controls. Informed consent had been previously obtained. The results were as follows:

1. Sons of alcoholics were about four times more likely to be alcoholic than were sons of nonalcoholics, whether raised by nonalcoholic foster parents or raised by their own biological parents. Sons of alcoholics raised by foster parents also were more likely to be divorced than were controls. Otherwise, sons of alcoholics and nonalcoholics did not differ with regard to a wide range of variables: depression, anxiety neurosis, personality disorders, criminality, drug abuse, or "heavy drinking" (defined as drinking daily and, on occasion, large amounts but without adverse consequences) (2).
2. Sons raised by their alcoholic parents differed from nonadopted controls only with regard to alcoholism, defined by operational criteria (2).
3. Of the adopted daughters of alcoholics, 2% were alcoholic and 2% more had serious problems from drinking. In the adopted control group, 4% were alcoholic. Of the nonadopted daughters, 3% were alcoholic and 2% were problem drinkers.

Between 0.1% and 1% of Danish women are alcoholic (14). Thus, in both the proband and control groups, a higher than expected prevalence of alcoholism was found. Nothing was known about the biological parents of the controls other than that they did not have a hospital diagnosis of alcoholism (the alcoholic parents of the probands were identified because they had been hospitalized with this diagnosis). Possibly some of the biological parents of the alcoholic controls were alcoholic. However, this could not be demonstrated one way or the other, and the findings from the daughter adoption study are inconclusive.

In both the adopted and nonadopted daughter groups there were low rates of heavy drinking. About 8% of the female subjects were heavy drinkers, as compared to nearly 40% of the male subjects. Therefore, of women who met the criteria for heavy drinking, a substantial number developed problems from drinking serious enough to require treatment.

Again, as with the male adoptees, the adopted-out daughters of alcoholics and controls did not differ with regard to other variables, such as depression or drug abuse. There has been speculation, based on family studies, that female relatives of alcoholics are apt to be depressed while male relatives are likely to suffer alcoholism (1). Indeed, 30% of the daughters raised by alcoholics, had been treated for depression by age 32, compared to about 5% of the controls. Apparently, growing up with an alcoholic parent increases the risk of depression in women but not in men, a susceptibility that does not exist if the daughters are raised by nonalcoholic foster parents.

This does not deny the possibility of a genetic predisposition to depression in female relatives of alcoholics, since many genetic disorders require an environmental "trigger" to become clinically apparent.

Summarizing the results of the Danish studies:

1. Children of alcoholics are particularly vulnerable to alcoholism whether raised by their alcoholic parents or by nonalcoholic foster parents.
2. The vulnerability is specific for alcoholism and does not involve increased risk for other psychopathology, including abuse of other substances.
3. Alcoholism is *not* on a continuum with "heavy drinking" or even "problem drinking" (defined as heavy drinking that results in problems but does not justify the term "alcoholism" as operationally defined in these studies).
4. More definitive conclusions could be drawn from studies of the sons of alcoholics than of the daughters because the female control adoptees also had a higher rate of alcoholism than would be anticipated from estimated prevalence in the general population. In general, heavy drinking was far less frequent in the women than in the men, but, of those women who did drink heavily, an unusually high percentage became alcoholic. As for an overlap between depression and alcoholism in the family histories, daughters of alcoholics raised by their alcoholic parents were more subject to depression than were controls, but this was not true of daughters raised by nonalcoholic foster parents. This suggests that in women environment may be more important in producing depression than in producing alcoholism.
5. The men in the study were relatively young to be diagnosed as alcoholic. Where they met the criteria for alcoholism, they almost always had received treatment. The women alcoholics had a severe form of alcoholism, requiring treatment, but they were somewhat older. Alcoholism often begins at a later age in women than in men (2), and in both men and women alcoholics future study may disclose an increase in prevalence. That is, the ratio of alcoholics to nonalcoholics in the four groups may possibly change with time.

Other Adoption Studies

Four other adoption studies have been conducted, two of them subsequent to the Danish studies.

In the early 1940s, Roe (15) obtained information about 49 foster children in the 20- to 40-year age group, 22 of normal parentage and 27 with a biological parent described as a "heavy drinker." Neither group had adult drinking problems. Roe concluded there was no evidence of hereditary influences on drinking.

This conclusion can be questioned on several grounds. First, the sample was small. There were only 21 men of "alcoholic" parentage and 11 of normal parentage. Second, the biological parents of the probands were described as "heavy drinkers," but it is not clear how many were alcoholic. Most had a history of antisocial behavior; none had been treated. All of the biological parents of the proband group in the Danish study received a hospital diagnosis of alcoholism, at a time when this diagnosis was rarely employed in Denmark.

In 1972, Schuckit et al (16) also studied a group of individuals reared apart from their biological parents who had either a biological parent or a "surrogate" parent with a drinking problem. The subjects were significantly more likely to have a drinking problem if their biological parent was considered alcoholic than if their surrogate parent was alcoholic.

More recently, Bohman (17) studied 2000 adoptees born between 1930 and 1949. He inspected official registers in Sweden for notations about alcohol abuse and criminal offenses in the adoptees and their biological and adoptive parents. There was a significant correlation between registrations for abuse of alcohol among biological parents and their adopted sons. Registered criminality in the biological parents did not predict criminality or alcoholism in the adopted sons.

Cadoret & Gath (18) studied 84 adult adoptees (18 years or older) separated at birth from their biological relatives and having no further contact with them. Alcoholism occurred more frequently in adoptees whose biological background included alcoholism than it did in other adoptees. Alcoholism did not correlate with any other biological parental diagnosis.

The above studies produced results similar to those found in the Danish adoption studies: alcoholism in the biological parents predicted alcoholism in their male offspring raised by unrelated adoptive parents but did not predict other psychiatric illness.

Two studies have been reported bearing on possible genetic modes of transmission. Kaij & Dock (19) tested the hypothesis of a sex-linked factor influencing the occurrence of alcoholism by comparing alcohol abuse rates in 136 sons of the sons vs 134 sons of the daughters of 75 alcoholics. No difference between the groups of grandsons was found, suggesting a sex-linked factor was not involved. Nearly half of the grandsons had alcohol problems.

Investigating the possibility that children of alcoholics metabolize alcohol abnormally, Utne and co-workers (unpublished data) compared the disappearance rate of blood alcohol in two groups of adoptees, ten with an alcoholic parent and ten without. There was no difference.

Familial Alcoholism

Studies reviewed here suggest that "familial alcoholism" may be a useful diagnostic category. At least for research purposes it might be useful to separate alcoholics into familial vs nonfamilial types.

Studies indicate that familial alcoholism should include these features:

1. A family history of alcoholism. If an alcoholic reports having one close relative who is alcoholic, he often reports having two or more.
2. Early onset of alcoholism. The sons of alcoholics in the Danish study were alcoholic by their late 20s. Alcoholism may begin in early life but usually male alcoholics are in their mid or late 30s before they are identified as alcoholics. In the Finnish twin study (4), younger identical twins were concordant for alcoholism more often than older twins. Jones (20) reports that younger alcoholics have alcoholic relatives more often than do older alcoholics.
3. Severe symptoms, requiring treatment at an early age. The alcoholic biological fathers in the Danish studies were identified because they were diagnosed in a Danish hospital. As noted, it is customary in many Danish hospitals to avoid the diagnosis of alcoholism when another diagnosis is available, e.g. a personality disorder. Therefore, it can safely be assumed that the alcoholic parents were severely alcoholic and this may explain why their offspring were so clearly alcoholic at a young age. In the study of Kaij (3), the concordance rate for alcoholism in identical twins rose as a function of the severity of the alcoholism. Another study (20) reported that "essential" alcoholism is associated with a family history of alcoholism more often than is "reactive" alcoholism. Essential alcoholism is defined as alcoholism apparently unrelated to external events (as "endogenous" depression is often contrasted with "reactive" depression); it connotes severity as well as lack of other psychopathology. Amark (21) noted that periodic (severe) alcoholics had a family history of alcoholism more often than did less severe alcoholics.
4. Absence of other conspicuous psychopathology. This was found in both the Danish studies (11–14) and the Bohman study (17).

While based on incomplete evidence, the concept of "familial alcoholism" may at least have research value.

Literature Cited

1. Winokur, G., Clayton, P. J. 1968. Family history studies: IV. Comparison of male and female alcoholics. *Q. J. Stud. Alcohol* 29:885
2. Goodwin, D. W. 1979. Alcoholism and heredity. *Arch. Gen. Psychiatry* 36:57–61
3. Kaij, L. 1960. *Studies on the etiology and sequels of abuse of alcohol,* thesis. Univ. Lund, Sweden
4. Partanen, J., Bruun, K., Markkanen, T. 1966. *Inheritance of Drinking Behavior: A Study on Intelligence, Personality, and Use of Alcohol of Adult Twins,* pp. 14–159. Helinski: Finn. Found. for Alcohol Stud.
5. Johnson, E., Nilsson, T. 1968. Alkoholkonsumption ho s monozygota och dizygota tvillingar. *Nord. Hyg. Tidskr.* 49:21
6. Vesell, E. S., Page, J. F., Passananti, G. T. 1971. Genetic and environmental factors affecting ethanol metabolism in man. *Clin. Pharmacol. Ther.* 12:192
7. Reed, T. E. 1977. Three heritable responses to alcohol in a heterogeneous

randomly mated mouse strain—inferences for humans. *Rutgers Univ. Cent. Alcohol Stud.* 38:618–32

8. Cruz-Coke, R., Varela, A. 1966. Inheritance of alcoholism. *Lancet* 2:1282
9. Fialkow, P. J., Thuline, M. C., Fenster, R. F. 1966. Lack of association between cirrhosis of the liver and the common types of color blindness. *N. Engl. J. Med.* 275:584
10. Goodwin, D. W., Schulsinger, F., Hermansen, L., et al. 1973. Alcohol problems in adoptees raised apart from alcoholic biological parents. *Arch. Gen. Psychiatry* 28:238–43
11. Goodwin, D. W., Schulsinger, F., Moller, N., et al. 1974. Drinking problems in adopted and nonadopted sons of alcoholics. *Arch. Gen. Psychiatry* 31:164–69
12. Goodwin, D. W., Schulsinger, F., Knop, J., et al. 1977. Alcoholism and depression in adopted-out daughters of alcoholics. *Arch. Gen. Psychiatry* 34:751–55
13. Goodwin, D. W., Schulsinger, F., Knop, J., et al. 1977. Psychopathology in adopted and nonadopted daughters of alcoholics. *Arch. Gen. Psychiatry* 34:1005–9
14. Goodwin, D. W. 1971. Is alcoholism hereditary? A review and critique. *Arch. Gen. Psychiatry* 25:545–49
15. Roe, A. 1944. The adult adjustment of children of alcholic parents raised in foster homes. *Q. J. Stud. Alcohol* 5:378–93
16. Schuckit, M. A., Goodwin, D. W., Winokur, G. 1972. A half-sibling study of alcoholism. *Am. J. Psychiatry* 128:1132–36
17. Bohman, M. 1978. Genetic aspects of alcoholism and criminality. *Arch. Gen. Psychiatry* 35:269–76
18. Cadoret, R., Gath, A. 1978. Inheritance of alcoholism in adoptees. *Br. J. Psychiatry* 132:252–58
19. Kaij, L., Dock, J. 1975. Grandsons of alcoholics. *Arch. Gen. Psychiatry* 32:1379–81
20. Jones, R. W. 1972. Alcoholism among relatives of alcoholic patients. *Q. J. Stud. Alcohol* 33:810
21. Amark, C. 1953. A study in alcoholism: clinical, social-psychiatric and genetic investigations. *Acta Psychiatr. Neurol. Scand.*, Suppl. 70, pp. 94–96

Ann. Rev. Med. 1981. 32:101–106
Copyright © 1981 by Annual Reviews Inc. All rights reserved

RETINAL VASCULAR OCCLUSION

❖7385

Walter H. Stern, M.D.

Department of Ophthalmology, University of California, San Francisco,
California 94143; and Veterans Administration Medical Center,
San Francisco, California 94121

Desmond B. Archer, F.R.C.S.

Department of Ophthalmology, Queens University of Belfast, Belfast, Ireland

Introduction

The retinal circulation is considered to be endarterial and a closed circuit since it is supplied by only one major afferent artery, the central retinal artery, and drained by only one efferent vessel, the central retinal vein. Knowledge of retinal vascular dynamics following occlusion of the major retinal vessels and their branches has expanded dramatically since the introduction of retinal fluorescein angiography, and has paved the way for important advances in the prognosis and therapy of a variety of retinal vascular occlusions.

Arterial Occlusions

The clinical picture of acute central retinal artery occlusion is characterized by sudden marked loss of vision in one eye. Ophthalmoscopic examination reveals narrowed arterioles and sludging of blood in the arterioles. The retina loses its transparency and turns milky white because of axonal swelling and necrosis of the inner retina. The characteristic "cherry red" spot in the fovea is due to the fact that the fovea does not contain retinal vessels and the thin retina in this area is nourished by the choroidal circulation, thus maintaining its transparency. Branch retinal artery occlusion produces similar ophthalmoscopic findings limited to the involved sector.

0066-4219/81/0401-0101$01.00

The etiology of central retinal artery occlusion is diverse; it includes atheroma formation or hemorrhage beneath the atheroma at the level of the lamina cribosa (1). Rarer causes include diseases such as temporal arteritis, periarteritis nodosa, sickle cell disease, trauma, and elevated intraocular pressure (2–4). Embolic phenomena appear to play a greater role in branch retinal occlusion than in central retinal artery occlusion. These emboli are usually atheromatous debris derived from the carotid arteries, e.g. cholesterol, platelets, and fibrin (5). Other sources of emboli are found in a variety of cardiac abnormalities including rheumatic vegetations, prolapsed mitral valves, and atrial myxomata (6, 7). There are many additional sources of emboli including self-injected retinal emboli (8).

Central retinal artery occlusion represents an ocular emergency. Clinical reports indicate that good retinal function may return after occlusion of the central retinal artery for up to one hour (4). Experimental studies indicate that retinal function may return after occlusion of the central retinal artery for up to two hours (9). One of the most useful therapeutic measures following acute central retinal artery occlusion is a rapid reduction in intraocular pressure by aspiration of a small amount of aqueous humor from the anterior chamber (4). Massage of the eye is also useful in lowering intraocular pressure or dislodging emboli. Vasodilation with retrobulbar injections of aminophylline or tolazoline or inhalation of 95% oxygen and 5% carbon dioxide has appeared to be useful in a few isolated cases. The use of anticoagulants and thrombolytic agents does not appear justified at the present time in view of their high risk and lack of demonstrable efficacy.

The medical workup of a patient with a central retinal artery occlusion should screen for those factors associated with atherosclerosis, including hypertension and a lipid profile. Additional tests for evidence of a collagen vascular disease should be performed. Any evidence of retinal emboli requires more extensive studies including cardiac ultrasonography and cardiovascular angiography. Black patients should always have a hemoglobin electrophoresis to screen for evidence of sickle cell disease.

Central Retinal Vein Occlusion

Central retinal vein occlusion can be divided into two distinct entities on the basis of clinical findings, pathogenesis, complications, and modes of therapy. These two entities are "venous stasis retinopathy" and "hemorrhagic retinopathy" (10).

Venous stasis retinopathy, also termed "impending" or "partial" central retinal vein occlusion, is encountered more frequently than hemorrhagic retinopathy. Patients often present with symptoms of blurred central vision that may improve during the day. Vision at the initial presentation is better than 20/50 in 90% of cases. Ophthalmoscopy in early cases reveals engorged tortuous retinal veins, scattered flame-shaped and punctate retinal hemorrhages, and optic disk hyperemia. Evidence of capillary occlusion such as cotton wool spots is rarely seen. The retinal arteries may be normal or slightly narrowed. Fluorescein angiography discloses marked venous stasis with leakage of fluorescein dye from the large veins and their tributaries, from the vessels on the optic nerve, and, in certain cases, from the perifoveal capillaries. Visual loss in this disease is due to the development of cystoid macular edema that results from leakage of fluid into the retina from the perifoveal capillaries. Certain features of fluorescein angiography such as a broken perifoveal capillary arcade and extensive capillary leakage, even in the absence of capillary closure, are useful in predicting the likelihood of cystoid macular edema (11).

Venous stasis retinopathy is self-limited and resolves in a period of months to several years. Most patients with good capillary perfusion and an intact perifoveal capillary net, as demonstrated on fluorescein angiography, will retain vision of 20/50 or better. If cystoid macular edema develops, then vision may decrease further. The disease affects both younger and older age groups. The pathogenesis in patients in their twenties and thirties is probably related to a phlebitis of the central retinal vein (12, 13). In older patients, it is probably secondary to atherosclerosis of the central retinal artery (10).

Hemorrhagic retinopathy is seen most frequently in older age groups. There is a marked deterioration of vision. Ophthalmoscopy reveals gross retinal hemorrhages, dilated and tortuous retinal veins, microaneurysms, cotton-wool spots, arteriolar narrowing, and diffuse retinal edema and optic nerve edema. Fluorescein angiography discloses marked capillary nonperfusion, venous stasis, and leakage of dye from the retinal veins and capillaries. Visual loss in this disease is due to retinal ischemia, intraretinal edema, and intraretinal hemorrhage. The most important complication is the development of iris neovascularization in approximately 50% of cases and the development of subsequent neovascular glaucoma in approximately 18% of patients (14).

The pathogenesis of hemorrhagic retinopathy is related to atherosclerosis and arteriosclerosis of the central retinal artery, intimal and endothelial cell proliferation in the central retinal vein, and compression of the central retinal vein within or just behind the lamina cribosa (15). Experimental studies by Hayreh (16) suggest that a temporary episode of retinal ischemia combined with closure of the central retinal vein is necessary to produce the picture of hemorrhagic retinopathy. This is refuted by Hamilton and colleagues (17), who feel that venous occlusion without concomitant arterial occlusion can produce progressive capillary closure and a picture resembling hemorrhagic retinopathy. Stagnation of blood flow has been involved in the pathogenesis of central retinal venous occlusion in a number of conditions including carotid disease, hyperviscosity syndromes, and glaucoma (18–20). Oral contraceptives have also been linked to central retinal vein occlusion (21).

Effective therapy appears to be limited in central retinal vein occlusion. Oral steroids are of limited value in venous stasis retinopathy in younger patients where one is presumably treating a papillophlebitis (10). Venous statis retinopathy is a self-limited disease in most cases, however, and no effective therapy for cystoid macular edema, which is the main complication of venous stasis retinopathy, has yet been demonstrated. Similarly, the efficacy of anticoagulation and fibrinolytic agents has not been clearly demonstrated (22, 23).

Panretinal photocoagulation is highly beneficial in preventing the secondary complication of neovascular glaucoma in the ischemic or hemorrhagic type of central retinal vein occlusion (14). Nonetheless, in hemorrhagic central vein occlusion, visual loss is severe because of extensive retinal ischemia, edema, and hemorrhage for which there is no effective therapy at present.

Branch Retinal Vein Occlusion

Branch retinal vein occlusion typically occurs at arteriovenous junctions where the vessels share a common inelastic adventitial coat. Sclerotic arteries, constrained at such sites, impinge directly on fellow veins; this causes venous stasis and, on occasion, thrombosis. The superotemporal fundus is the most common location for tributary vein occlusion, as more arteriovenous crossings exist here than in any other quadrant. Vein occlusions are particularly common in hypertensive individuals and in patients with vascular changes brought on by advancing age. A variety of other etiologic factors exist, many of which were mentioned in the pathogenesis of central retinal vein occlusion.

The sectorial ophthalmoscopic and fluorescein angiographic features of branch vein occlusion are similar to those found in central vein obstruction except for two important features: retinal capillary collaterals and retinal and optic nerve neovascularization.

Capillary collaterals follow almost immediately after the acute occlusion and may be identified on fluorescein angiography within a few days of the obstruction. Some of these collaterals may give way under the excessive hemodynamic load and fine hemorrhages occur in the immediate microvasculature of the obstructed vein. Intraretinal hemorrhages clear rapidly once collaterals develop. Established collaterals are clinically observable within a few weeks and persist for many years or indefinitely.

Retinal neovascularization in the occluded quadrant is related to the extent of ischemia noted on fluorescein angiography (24). The overall incidence of retinal neovascularization is approximately 25%. When only patients with marked ischemia, noted on fluorescein angiography, are included, the risk of neovascularization rises to 62% (25, 26).

Most studies of the natural course of branch retinal vein obstruction show that about 50% of patients regain normal or near normal vision (20/20 to 20/30) and that the retinal microvasculature adjusts in large part to the venous obstruction (27). Vessels that fail to regain competence within the first few months of the acute obstruction generally remain permeable to fluorescein, and, when the macula is involved, the vision usually deteriorates slowly as secondary degenerative changes occur.

The management of branch vein occlusion is controversial in view of the generally good prognosis for visual acuity, as noted in studies of the natural history of the disease. Anticoagulants, carbogen (oxygen and carbon dioxide) inhalations, fibrinolysin, and other medical therapies have been recommended for patients with branch vein occlusion. However, none of these have been shown to improve significantly the natural course of the disease. If venous occlusion is secondary to vasculitis, systemic corticosteroids may have a place in therapy.

Fluorescein angiography is important in predicting the likelihood of a patient developing retinal and optic nerve neovascularization. Scatter photocoagulation to the involved retinal quadrant is extremely effective in causing involution of the retinal and disk new vessels (24).

Photocoagulation has also been used to treat macular edema, although controlled trials to date have not shown a significant difference between untreated eyes with established edema in terms of improved central vision (26). Retinal holes or tears associated with preretinal neovascularization and inner retinal ischemia have been effectively treated by photocoagulation if flat, or by conventional retinal detachment surgery if elevated (28). Where a severe vitreous hemorrhage complicates a tributary vein occlusion, pars plana vitrectomy is useful in removing the vitreous blood and clearing the visual axis.

Conclusion

Retinal arterial and venous occlusions continue to pose difficult problems for the clinician in terms of effective therapy. The most important advances in the last two decades have been the use of fluorescein angiography as a guide to visual prognosis in vein occlusions and the use of retinal photocoagulation to destroy or cause regression of retinal and iris neovascularization. Effective treatment of the systemic and ocular conditions responsible for these vascular occlusions will continue to be the most important task for the clinician.

Literature Cited

1. Dahrling, B. E. 1965. The histopathology of early central retinal artery occlusion. *Arch. Ophthalmol.* 73: 506–10
2. Wagener, H. P., Hollenhorst, R. W. 1957. The ocular lesions of temporal arteritis. *Trans. Am. Ophthalmol. Soc.* 55:249–73
3. Bernstein, A. 1935. Periarteritis nodosa without peripheral nodules diagnosed antemortem. *Am. J. Med. Sci.* 19:317–25
4. Radius, R. L., Finkelstein, D. 1976. Central retinal artery occlusion (reversible) in sickle trait with glaucoma. *Br. J. Ophthalmol.* 60:428–30
5. Ball, C. J. 1966. Atheromatous embolism to the brain, retina and choroid. *Arch. Ophthalmol.* 76:690–95
6. Penner, R., Font, R. L. 1969. Retinal embolism from calcified vegetations of aortic valve: a spontaneous complication of rheumatic heart disease. *Arch. Ophthalmol.* 81:565–68
7. Anderson, J. D., Lubow, M. 1973. Atrial myxoma as a source of retinal embolism. *Am. J. Ophthalmol.* 76: 769–72
8. Schatz, H., Drake, M. 1979. Self-injected retinal emboli. *Ophthalmology* 86:468–83
9. Hamasaki, D. J., Kroll, A. J. 1968. Experimental central retinal artery occlusion. An electrophysiological study. *Arch. Ophthalmol.* 80:243–48
10. Hayreh, S. S. 1976. So-called "central retinal vein occlusion." I. Pathogenesis, terminology, clinical features. *Ophthalmologica* 172:1–13
11. Laatikainen, L., Kohner, E. M. 1976. Fluorescein angiography and its prognostic significance in central retinal vein occlusion. *Br. J. Ophthalmol.* 60: 411–18
12. Lyle, K. T., Wybar, K. 1961. Retinal vasculitis. *Br. J. Ophthalmol.* 45: 778–88
13. Lonn, L. I., Hoyt, W. F. 1966. Papillophlebitis. A case of protracted yet benign optic disc edema. *Eye, Ear, Nose, Throat Mon.* 45:62–68
14. Laatikainen, L., Kohner, E. M., Khoury, D., Blach, R. K. 1977. Panretinal photocoagulation in central retinal vein occlusion: a randomised controlled clinical study. *Br. J. Ophthalmol.* 61:741–53
15. Klien, B. A., Olwin, J. H. 1956. A survey of the pathogenesis of retinal venous occlusion. *Arch. Ophthalmol.* 56: 207–47
16. Hayreh, S. S. 1978. Experimental retinal vascular occlusion. I. Pathogenesis of central retinal vein occlusion. *Arch. Ophthalmol.* 96:311–23
17. Hamilton, A. M., Kohner, E. M., Rosen, D., Bird, A. C., Dollery, C. T. 1976. Experimental retinal branch vein occlusion in rhesus monkeys. I. Clinical appearances. *Br. J. Ophthalmol.* 63: 377–87
18. Kearns, T. P., Hollenhorst, R. W. 1963. Venous stasis retinopathy of occlusive disease of the carotid artery. *Proc. Staff Meet. Mayo Clin.* 38:304–12
19. Rothstein, T. 1972. Bilateral, central retinal vein closure as the initial manifestation of polycythemia. *Am. J. Ophthalmol.* 74:256–60
20. Bertelson, T. J. 1961. The relationship between thrombosis in the retinal veins and primary glaucoma. *Acta Ophthalmol.* 39:603–13
21. Stowe, C. G. III., Zakov, N. Z., Albert, D. M. 1978. Central retinal vascular occlusions associated with oral contraceptives. *Am. J. Ophthalmol.* 86:798–801
22. Vannas, S., Raitta, C. 1966. Anticoagulant treatment of retinal venous occlusion. *Am. J. Ophthalmol.* 62:874–84
23. Kohner, E. M., Hamilton, A. M., Bulpitt, C. J., Dollery, C. T. 1974. Streptokinase in the treatment of central retinal vein occlusion. *Trans. Ophthalmol. Soc. UK* 94:599–603
24. Archer, D. B., Ernest, J. T., Newell, F. W. 1974. Classification of branch retinal vein obstruction. *Trans. Am. Acad. Ophthalmol. Otol.* 78:148–65
25. Krill, A. E., Archer, D. B., Newell, F. W. 1971. Photocoagulation in complications secondary to branch vein occlusion. *Arch. Ophthalmol.* 85:48–60
26. Shilling, J. S., Kohner, E. M. 1976. New vessel formation in retinal branch vein occlusion. *Br. J. Ophthalmol.* 60: 810–15
27. Michels, R. G., Gass, J. D. M. 1974. The natural course of retinal branch vein obstruction. *Trans. Am. Acad. Ophthalmol. Otol.* 78:166–77
28. Gutman, F. A., Zegarra, H. 1976. Retinal detachment secondary to retinal branch vein occlusions. *Trans. Am. Acad. Ophthalmol. Otol.* 81:491–96

Ann. Rev. Med. 1981. 32:107–22
Copyright © 1981 by Annual Reviews Inc. All rights reserved

HOST DETERMINANTS IN STAPHYLOCOCCAL DISEASE

❖7386

Jan Verhoef, M.D. and Henri A. Verbrugh, M.D.

Laboratory of Microbiology, University of Utrecht, Catharijnesingel 59,
3511 GG Utrecht, The Netherlands

INTRODUCTION

Despite impressive gains in knowledge about the microbe and the host, the understanding and control of infections caused by staphylococci are still far from complete. *Staphylococcus aureus* continues to cause serious infectious diseases with a relative high morbidity and mortality. These are classical examples of opportunistic infections: deviations from the state in which the host controls bacterial colonization and prevents tissue invasion.

The staphylococcus possesses numerous virulence factors that can act independently to induce a variety of symptoms and lesions. Therefore, clinical disease caused by staphylococci may vary widely. Although staphylococci have a special preference for invasion of soft tissues and bones, every organ system can be affected (1, 2).

This paper reviews progress made in understanding the interaction between virulence factors of staphylococci and defense mechanisms of the host as they relate to colonization and invasion. In addition, defects in host resistance are discussed, as these may indicate the relative importance of host determinants in staphylococcal disease. The potential virulence factors of staphylococci and their effects on host defense are summarized in Table 1.

COLONIZATION

In order to successfully parasitize the host, microorganisms must first firmly attach to host cells. This phenomenon of attachment is called bacterial adherence: a specific and selective binding of certain bacterial surface components to receptors in the membrane of host cells. Bacterial adherence

0066-4219/81/0401-0107$01.00

Table 1 Potential virulence factors of *Staphylococcus aureus*

Virulence factors	Proposed effect on host defense
Cell surface components	
Peptidoglycan	See Table 2
Protein A	Interference with opsonization
Capsular material	Interference with opsonization
Extracellular products	
Coagulase	Inhibition of phagocytosis (?)
Catalase	Inhibition of intraleukocytic killing
Enterotoxin	Gastrointestinal disturbance
Epidermolytic toxin	Exfoliative skin lesions
Leukocidin	Cytotoxic, leukotoxic
α-Hemolysin	Leukotoxic, induces dermonecrosis
Ability to survive within phagocytic cells	

is generally thought to be an important first step in the pathogenesis of infections (3). But surprisingly, little is known about the factors that control staphylococcal adherence.

Staphylococci can adhere to many different cell types (4–6) and readily colonize both skin and mucous membranes. The carriage rate for *S. aureus* on moist areas of the skin and in the anterior nares of healthy people varies from 10–40% (7). In a number of conditions, the *S. aureus* carriage rate is even higher. Over 60% of patients on long-term hemodialysis harbor *S. aureus* (8). In these patients, the carriage rate increases with time on hemodialysis but does not correlate with the frequency of hospital admission. A higher incidence of staphylococcal bacteremia was found in carriers compared to patients who were not colonized. Insulin-injecting diabetics and parenteral drug abusers have higher rates of staphylococcal carriage (9, 10). Also of interest, staff members of the dialysis unit as well as patients with chronic renal failure without regular hemodialysis have an increased carriage rate, while diabetics taking oral hypoglycemic agents and former drug abusers do not (8–10).

Adherence

Frost and co-workers (4, 5) demonstrated selective adherence of *S. aureus* to isolated epithelial cells from the bovine mammary duct. *Streptococcus agalactiae* adhered equally well. The adhesive properties of the bacteria paralleled their prevalence in bovine mastitis. Only certain types of cells, those not containing microvilli, were found to attach bacteria (5). Similar observations on *S. aureus* adherence were reported by Aly and co-workers (6, 11). Their studies showed that *S. aureus, S. epidermidis, S. pyogenes, Pseudomonas aeruginosa,* and diphtheroids firmly attached to washed nasal cells. *Klebsiella pneumoniae* and *Streptococcus viridans* did not show a specific adherence. An increased number of staphylococci adhered to nasal mucosal cells from *S. aureus* carriers compared to cells from noncarriers (6), and mucosal cells from adults showed greater numbers of bacteria attached than did cells obtained from neonates. Interestingly, a virulent staphylococcal strain showed better adhesive properties than did the relatively avirulent 502 A strain of *S. aureus* (11). Heat killing of staphylococci only marginally reduced their capacity to adhere (6). However, pretreatment of staphylococci with specific antisera completely blocked staphylococcal adherence (5). Fainstein & Musher (12) demonstrated that pharyngeal cells from patients with acute respiratory illness and those from volunteers experimentally infected with influenza A vaccine virus showed increased adherence of *S. aureus.* They postulated that alteration in mucosal cells leading to increased bacterial adherence may play a role in the pathogenesis of superinfection by these microorganisms in patients with viral respiratory disease. Although teichoic acid has recently been proposed as one of the *S. aureus* cell wall components important for adherence (13), most of the adherence factors from both the host and the parasite are largely unknown.

Interference

The presence of the regular microbial residents makes it more difficult for other microorganisms to become an established part of the microflora. The ability of bacteria to interfere with the multiplication of other bacteria has been demonstrated between numerous microorganisms, including staphylococci (14). Interference between strains of *S. aureus* was used successfully as a tool in curtailing epidemics of staphylococcal disease in nurseries and in preventing recurrent furunculosis in patients (15, 16). In the latter case, it is essential to eradicate the pathogenic staphylococcus first, before the interfering strain, usually *S. aureus* 502 A, "takes" and persists (14). Because minor skin infections and, rarely, more serious abscesses and one case of fatal septicemia have been reported (17), routine use of this strain in nonepidemic situations or in patients with suspected or proven defects in host defenses was discouraged. The basis of bacterial interference among staphylococci is poorly understood, and theoretically, may involve nutrient competition, release of toxic metabolites or bacteriocins, and competition for adherence sites on host cells (18, 19). The commensal flora is also influenced by the pH of the skin and mucous membrane and by antibiotics, tissue macrophages, and neutrophils (20).

INVASION

Once the skin and mucosal membranes are breached, the staphylococci may become highly pathogenic invaders. *S. aureus* is the most frequent bacterial isolate in wound infections (21), cutaneous abscesses (22), deep tissue abscesses secondary to nonpenetrating trauma (23), shunt infections and bacteremia in hemodialysis patients (7), pyogenic infections of the spine (24), blood cultures of patients with bacteremia (25, 26), and in acute bacterial endocarditis (27).

Local defense mechanisms can be damaged, not only by trauma, but also by virus infections. Staphylococcal pneumonia is well known as a life-threatening complication of influenza virus respiratory tract infection (28, 29). At least three distinctive mechanisms for impairment of antibacterial pulmonary defenses by viruses have been proposed.

First, influenza virus produces extensive desquamation of bronchial epithelium and areas of consolidation in the lung. These lesions prevent bacterial clearance from the lungs by the mucociliary transport mechanism; consequently the inhaled staphylococci remain localized and multiply in the infected areas (30).

Second, the adherence of *S. aureus* to mucosal epithelial cells may be increased during respiratory viral infections (12). Austin & Daniels (31) showed that staphylococcal strains that contained protein A had an increased adherence to influenza-infected kidney cell monolayers pretreated with anti-influenza serum; strains not containing protein A did not. They concluded that protein A mediated adherence through its ability to react nonspecifically with the Fc portion of mammalian IgG. Protein A is detectable in 99.5% of *S. aureus* strains but not in *S. epidermidis* strains (32–35).

A third possible mechanism is the impaired phagocytic function of alveolar macrophages. These cells play a crucial role in destroying inhaled staphylococci (36). Jakab et al (37) showed that virus-induced impairment of alveolar macrophage killing was related to a defect in intracellular fusion between the lysosomes, which contain the bactericidal agents, and the phagosome, which contains the ingested microorganisms. These findings correlate well with clinical and experimental observations on leukocyte function during influenza virus infection in man. Defective phagocytic function of polymorphonuclear and mononuclear cells (38, 39) has been reported in patients with acute influenza respiratory infection. Recent work in our laboratory demonstrated that influenza virus induces a respiratory burst in human peripheral blood neutrophils and monocytes (unpublished data). Thus, the capacity of phagocytes to produce bactericidal oxygen products may be depleted by viral infection; through auto-oxidation, toxic metabolites may accumulate causing decreased phagocyte function and accelerated cell death. Additional information relevant to the host response of the human respiratory system may now be obtained by use of fiberoptic bronchoscopy to retrieve bronchoalveolar macrophages and alveolar lining material of man.

PHAGOCYTOSIS AND DESTRUCTION

Once the outer membranes are disrupted, ingestion and killing by phagocytic cells become the major line of defense against infection. Phagocytosis is the process whereby cells transport particulate objects from the extracellular environment into intracellular vacuoles (phagosomes). Although many types of cells can be induced to engulf particles, the "professional" phagocytes are the neutrophilic polymorphonuclear leukocytes and the mononuclear phagocytes, i.e. the blood monocyte and the tissue macrophage (for reviews, see 40–43). These cells have special capabilities to ingest and kill particles. Phagocytosis is preceded by direct migration of the phagocytic cell toward the focus of infection (chemotaxis). A number of separate steps can be distinguished during phagocytosis of *S. aureus*, including (*a*) alteration of the bacterial surface through opsonization, (*b*) attachment and ingestion of the opsonized bacteria to membrane receptors of phagocytes, and (*c*) intracellular killing and degradation of the staphylococci. This third step involves a multiplicity of mechanisms, all of which are centered around two cellular events: degranulation, i.e. fusion between the phagosome and granules present in the cytoplasm of the phagocytes; and the initiation of the so-called respiratory burst, whose function is to produce a group of highly reactive microbicidal agents by the partial reduction of oxygen.

Opsonization

Recruitment of adequate numbers of phagocytes to the infection site, as well as recognition of the staphylococci by these cells, is essential for the eradication of invading organisms. Recognition of staphylococci by phagocytes is mediated by opsonins, which are IgG antibodies and factors from the heat-labile complement system (44–46). These opsonins become firmly fixed to surface structures of the microbe and are recognized by the leukocyte. Receptors with structural identity have not yet been defined, but the phagocyte functions as though it possesses membrane receptors specific for these opsonins.

Although staphylococci are adequately opsonized in immune sera containing high levels of specific antibody without the participation of complement, the heat-labile complement system provides the majority of opsonic molecules for staphylococcal phagocytosis in healthy, nonimmune individuals (46–48). Considerable attention has been focused on the nature of the opsonic antibody molecule, and on the opsonic function of the complement system.

Antibody directed against the teichoic acid (49) and peptidoglycan (50, 51) moieties of the cell wall may be of primary importance in staphylococcal opsonization. Recent studies demonstrate that cell wall peptidoglycan can absorb the heat-stable opsonic activity from normal and immune human sera, whereas α or β teichoic acids cannot (45, 50, 51). In addition, Karakawa & Young (51) and Verbrugh et al (52) were able to demonstrate a direct opsonic effect of peptidoglycan-specific antibodies isolated from hyperimmune rabbit sera. At least three antigenic sites can be distinguished in bacterial peptidoglycans (53): (*a*) glycan strands that are chemically rather uniform among the bacterial species with *N*-acetyl-glucosamine as the immunodominant group, (*b*) non-cross-linked pentapeptide with COOH-terminal D-alanyl-D-alanine as the immunodominant group, and (*c*) interpeptide bridges that are hexa- or pentaglycines. Antibodies directed against these interpeptide bridges are highly specific and have been used to distinguish staphylococci from micrococci (54).

The immunological specificity of the opsonic peptidoglycan antibodies for staphylococcal phagocytosis is uncertain. Karakawa et al (51) found that peptidoglycan antibodies raised against group-A-variant streptococci stimulated staphylococcal uptake by human neutrophils. Recently, these investigators also showed that antibodies against *Escherichia coli* peptidoglycan were able to opsonize *S. aureus*, and they suggested a D-alanyl-D-alanine specificity for the opsonic immunoglobulins (55). In contrast, Humphreys et al (45) could not absorb peptidoglycan antibodies that were opsonic for *S. aureus* with purified peptidoglycans from *S. epidermidis* and *S. thermophilus*. The use of (monoclonal) antibodies with homogenous specificity against each of the three immunodominant sites of the peptidoglycan matrix is needed to establish the nature of the opsonic, peptidoglycan-directed antibody molecule.

In nonimmune serum, staphylococcal opsonization is determined by the capacity of the various cell wall components to activate the complement system, which leads to deposition of opsonic C3b on the bacterial surface. The binding of human C3 to *S. aureus* during opsonization can be quantitatively measured with radiolabeled C3 (56) or with fluorescence spectrofluorometry (57, 58). Using this latter method, Verbrugh et al (57) demonstrated a direct relationship between the rate of staphylococcal C3 fixation and the rate of opsonization in nonimmune human serum. All major cell wall components of *S. aureus* activate complement in normal human serum (58–62), and may thus contribute to complement-mediated staphylococcal opsonization. In the absence of antibody or a functional classical complement pathway, staphylococcal opsonization may proceed by activation of only the alternative complement pathway (46, 63), which is mediated exclusively by peptidoglycan. From these studies, it is now clear that the peptidoglycan matrix of the cell wall is of great importance in the opsonization of the staphylococci by heat-stable as well as heat-labile serum opsonins (60).

Chemotaxis

The directional movement of phagocytic cells into the site of infection (chemotaxis) is one of the initial host responses to staphylococcal invasion (for review, see 40). The migration of the cells occurs under the influence of soluble mediators (chemotactic factors). Some extracellular factors produced by staphylococci act as these chemotaxins (64, 65). The precise nature of these factors is unknown. However, the major neutrophil chemotactic response to staphylococci results through activation of the complement system (66). By virtue of their ability to activate complement, all major cell wall components of *S. aureus* contribute to the generation of chemotaxins. In this way, they act as chemotaxigens. Important chemotactic factors are a heat-stable cleavage product of the fifth component of complement (C5a), a heat-labile product of C3 (C3a), C3 convertase, and the trimolecular complex C567 (40). These factors are generated not only when complement is activated, but also after interaction with bacterial proteases, plasmin, trypsin, etc.

Attachment and Ingestion

Once opsonized, staphylococci are readily phagocytized by human polymorphonuclear phagocytes. The phagocyte functions as though it possesses membrane receptors specific for opsonins. Different roles have been suggested for neutrophil IgG and complement receptors in the attachment and the ingestion phases of phagocytosis. Some investigators (67) find that complement receptors of macrophages and neutrophils can only mediate attachment but cannot induce ingestion of coated sheep erythrocytes or immune complexes. Fc receptors, however, were able to mediate both attachment and ingestion of opsonized particles. In contrast, the uptake of *S. aureus,* opsonized in either diluted normal serum or immune-globulin-deficient sera occurs via trypsin-sensitive membrane receptors, presumably C3b receptors, of the leukocytes (68).

The interaction between opsonins and the receptors results in the formation of pseudopods, which ultimately engulf the staphylococcus (ingestion) (see 40 for details).

Killing and Destruction

After being phagocytized most intracellular staphlococci are rapidly killed and degraded within the phagocytic vacuoles. However, some bacteria may survive within the leukocyte for prolonged periods of time. These survivors may outlive the phagocytic cells, which could explain the chronicity of some staphylococcal infections (69).

The intracellular killing of staphylococci consists of oxygen-dependent and oxygen-independent bactericidal mechanisms. Leukocytes from healthy donors respond to phagocytosis by increasing oxygen uptake, and shifting the metabolism to the hexose monophosphate shunt. Hydrogen peroxide and highly staphylocidal oxygen- and hydroxyl-radicals are produced (oxygen-dependent killing system). Oxygen-independent staphylocidal systems in the phagocytes include the low pH within the vacuole, lysozyme, lactoferrin, and granular cationic proteins (40). Staphylococci are sensitive to lactic acid at pH 4.0–5.5, to lactoferrin, and to cationic proteins. The increased virulence of staphylococci grown in vivo as compared to strains grown in vitro may in part be due to their increased resistance to intraleukocytic killing by cationic proteins (70).

HUMORAL IMMUNITY

Antibacterial and Antitoxic Antibodies

During a staphylococcal infection, antibodies are produced against extracellular staphylococcal products and against cell wall antigens. In animal experiments, a clear distinction can be made between bactericidal, antibacterial, and antitoxic immunity to *S. aureus.* Koenig et al (71) found that mice immunized with staphylococcal toxoid were completely resistant to staphylococcal α toxin (a membrane-damaging toxin), but were not protected against intraperitoneal challenge with live staphylococci. On the other hand, antibacterial immunity, obtained by intravenous injection of heat-killed bacteria, did not protect mice from the lethal effect of α toxin.

From the considerable literature dealing with the various aspects of antitoxic and antibacterial immunity (see 72 for details), it is clear that no single toxin or cell wall structure has yet been demonstrated to be the all-important protection-inducting antigen. Also, attempts to correlate serum antitoxin titers to the resistance, presence, and activity of staphylococcal disease have met with mixed results (72). The importance of antibody against extracellular products seems to be limited to altering the course of already established staphylococcal infections. The levels of heat-stable, opsonic antibodies can be enhanced by either experimental infection or immunization, or during natural infection in animals and man (73). Also antibodies against teichoic acid are produced. This has led to the development of teichoic acid antibody assays as a clinical tool in the diagnosis of staphylococcal infection (74).

CELL-MEDIATED IMMUNITY

A cellular immune response against staphylococci, as evidenced by delayed hypersensitivity, has been demonstrated in immunized mice (75) and in patients with various types of recurrent *S. aureus* infections (76). The induction of delayed hypersensitivity appears to depend upon multiple or prolonged encounters with the microorganism and requires repeated injections of viable cocci in animals. This type of acquired immunity against staphylococci is transferable with leukocytes but not with serum. Repeated infections of rabbits and mice with staphylococci, followed by an appropriate elicitation step, increases the staphylocidal capacity of the peritoneal macrophages (77).

However, harmful effects of acquired cell-mediated immunity have repeatedly been demonstrated in systemic and local infections. Easmon & Glynn (78) observed that the incidence and extent of necrotic skin lesions was greatest in mice given lymph node lymphocytes from sensitized donors. In contrast, T-lymphocyte-deficient nude mice, as well as T-cell-depleted mice showed mild staphylococcal lesions with no necrosis. These results also indicated that immune serum may override and prevent this deleterious effect of cell-mediated immunity. Splenic lymphocytes from mice displaying delayed hypersensitivity to *Staphylococcus aureus* release macrophage-activating lymphokines when cultured in vitro in the presence of heat-killed bacteria (79). Both cell wall peptidoglycan and *S. aureus* membrane proteins have been thought (78, 80) to contain the major determinants for delayed hypersensitivity. No cross-reactivity was observed when antigens from gram-negative bacteria were used. The role of cellular immunity in human defense against *S. aureus* remains to be clearly defined and is therefore a promising field for laboratory and clinical investigations.

INTERFERENCE OF STAPHYLOCOCCAL PRODUCTS WITH HOST DEFENSE

Extracellular Products

Leukocidin and α and β toxin produced by *S. aureus* inhibit chemotaxis of human blood neutrophils and monocytes (65). On the other hand, some extracellular products act like chemotaxins and stimulate migration of phagocytes, as was noted before (64, 65). The potential toxic effects of many extracellular staphylococcal products on the tissues are reviewed extensively (81) and are beyond the scope of this discussion.

Cell Wall Components

Some *S. aureus* strains possess acidic polysaccharide capsules or pseudocapsules. These encapsulated strains are not effectively opsonized in normal serum (82); complement is not an adequate opsonin for these strains. Antibodies specific for the capsular material are essential for efficient opsonization and subsequent phagocytosis. Opsonization can also be inhibited by protein A (83). As was mentioned before, protein A can react with the Fc fragment of IgG, preventing binding of the opsonic antibody molecule with the so-called Fc receptor in the membrane of the phagocyte. Also, extracellular protein A, which is produced during bacterial growth, can consume complement and therefore decrease the amount of complement available for opsonization.

Recent attention has focused on the "endotoxic" potential of peptidoglycan (84, 85), which may be critical in determining the severity of the infection. Although many of the toxic effects were first observed with peptidoglycans from streptococci, the structurally related staphylococcal peptidoglycan behaves very similarly. Thus, staphylococcal peptidoglycan is pyrogenic when injected intravenously into rabbits (84). Intradermal injection of purified peptidoglycan induces inflammatory skin lesion in a dose-dependent fashion (84). Hill (86) observed that the resistance against intracutaneous challenge with live staphylococci is greatly reduced by prior injection of peptidoglycan into the challenge site. Peptidoglycan inhibited the early, beneficial stage of fluid exudation at the site of infection, and was called "aggressin." Easmon et al (87) suggested that the aggressin activity was due to the ability of peptidoglycan to prevent the release of kinins by interfering with the Hageman factor–kinin pathway. In addition, peptidoglycan reduces leukocyte chemotactic responsiveness (88), and this effect, which is neutralized by peptidoglycan-specific antibody (89), may also contribute to the increased severity of the lesions. On the other hand, staphylococcal peptidoglycan may also exert some beneficial effects on the host that makes the study of its role in health and disease even more challenging (see Table 2).

The capacity of staphylococci to consume human complement may also have serious side effects. Lew et al (90) recently reported that at the site of infection complement consumption by bacteria may be extensive, resulting in release of appreciable quantities of complement breakdown products and low levels of remaining opsonic activity. In addition to lowering of locally

Table 2 Properties of staphylococcal peptidoglycan

Property	Effect
Pyrogenicity	Increases nonspecific resistance to bacterial infection
Shwartzman reactivity	Mitogenicity
Induction of inflammatory skin lesions and lesions in internal organs	Adjuvant activity
Agressin activity	Macrophage activation
Immunosuppressive activity	Anti-tumor activity ?
Complement activation (both pathways)	Immunogenicity (B and T cell systems)

available heat-labile opsonins, breakdown products of complement may have a direct deleterious effect on phagocytic cells. These products, primarily C5a, may lead to aggregation and immobilization of granulocytes at a site distant from the focus of infection (91).

DEFECTS IN HOST RESISTANCE

Defects in the host resistance that lead to an increased susceptibility to infections are extensively reviewed (40, 92, 93). Several clinical entities have been described in which a defective staphylocidal function of leukocytes has been found. The first syndrome in which the pathogenesis of disease was clearly associated with defective bactericidal function of polymorphonuclear leukocytes was chronic granulomatous disease (94). Phagocytes from patients with this congenital disease do not respond normally to phagocytosis with an increase in oxygen uptake. There is little shift in metabolism to the hexosemonophosphate shunt, and hydrogen peroxide does not accumulate in the phagocytic vacuoles. Microorganisms that produce their own hydrogen peroxide, however, do not survive within the intraleukocyte environment. *S. aureus* produces catalase, an enzyme that breaks down the hydrogen peroxide produced by the staphylococcus, and is not killed by its own peroxide. *S. aureus* is thus well equipped for prolonged survival in such leukocytes with defective hydrogen-peroxide-producing capacity.

Patients with hematologic malignancies and patients receiving cytotoxic chemotherapy may have both qualitative and quantitative leukocyte defects. Staphylococcal infections are serious and recurrent in such patients (for reviews see 40, 95). Recurrent staphylococcal infections are also seen in patients with diabetes mellitus and renal insufficiency. Leukocyte dysfunction may account for this propensity for infection; defects have been described in chemotaxis, phagocytosis and bactericidal activity (40). A number of conditions with defective leukotaxis have been reported and recurrent staphylococcal infections have been associated with such defects (40). One such condition is called Job's syndrome, a condition that is also associated with greatly elevated serum IgE levels. The decreased leukotactic responsiveness of the patients with high IgE levels is not related to a direct effect of the high concentrations of IgE, since the serum from these patients does not impair the chemotactic activity of control leukocytes. It is postulated that histamine and histamine-induced cyclic adenosine monophosphate accumulate within the neutrophils of these patients and depress chemotactic responsiveness. The high levels of IgE contain antistaphylococcal antibodies (96).

Abnormal phagocytosis due to a deficiency in the heat-labile opsonic activity of serum may also lead to serious infections. Some of the patients with deficiencies in components of the complement system suffer from recurrent staphylococcal infections. Also, infections are frequently seen in patients with a diminished heat-stable opsonic activity such as in patients with agammaglobulinemia and with certain lymphoreticular malignancies.

CONCLUSION AND PERSPECTIVE

Considering the armament of the staphylococcus, i.e. its exotoxins, the antiphagocytic cell wall components, its peptidoglycan with "endotoxic" activity, and its capacity to survive the intraphagocytic environment, it is not surprising that *Staphylococcus aureus* remains a serious cause of infection. Chemotherapy has changed the clinical spectrum but has not eradicated serious staphylococcal infections. The pathogenesis of staphylococcal disease remains poorly understood. Investigations of normal cellular and humoral responses to staphylococci have provided some insight into the significant microbial and host factors. The study of these responses in patients with defective immunological mechanisms has been especially valuable in increasing our knowledge of host response to this microorganism.

Future advances in the prevention and therapy of staphylococcal disease will depend on advances in knowledge of the molecular biology of the microbe and the mechanisms of host defense. Attention should be especially focused on the host factors that control colonization and on microbial and host factors responsible for invasion and generalized infection.

ACKNOWLEDGMENT

We thank Dr. D. M. Musher for reviewing the manuscript critically.

Literature Cited

1. Musher, D. M., McKenzie, S. O. 1977. Infections due to *Staphylococcus aureus. Medicine* (Baltimore) 56:383–409
2. Beaty, H. N. 1980. *Staphylococcus aureus* bacteremia. In *Current Clinical Topics in Infectious Diseases,* ed. J. S. Remington, M. N. Swartz, 1:34–44. New York: McGraw-Hill, 331. pp.
3. Reed, N. W., Williams, R. C. 1978. Bacterial adherence: first step in the pathogenesis of certain infections. *J. Chron. Dis.* 31:67–92
4. Frost, A. J. 1975. Selective adhesion of microorganisms to the ductular epithelium of the bovine mammary gland. *Infect. Immun.* 12:1154–56
5. Frost, A. J., Wanasinghe, P. D., Woolcock, J. B. 1977. Some factors affecting selective adherence of microorganisms in the bovine mammary gland. *Infect. Immun.* 15:245–53
6. Aly, R., Shinefield, H. I., Strauss, W. G., Maibach, H. I. 1977. Bacterial adherence to nasal mucosal cells. *Infect. Immun.* 17:546–49
7. Williams, R. E. O. 1963. Healthy carriage of *Staphylococcus aureus:* its prevalence and importance. *Bacteriol. Rev.* 27:56–71
8. Kirmani, N., Tuazon, C. U., Murray, H. W., Parrish, A. E., Sheagren, J. N. 1978. *Staphylococcus aureus* carriage rate of patients receiving long-term hemodialysis. *Arch. Intern. Med.* 138: 1657–59
9. Tuazon, C. U., Perez, A., Kishaba, T., Sheagren, J. N. 1975. *Staphylococcus aureus* among insulin-injecting diabetic

patients. An increased carriage rate. *J. Am. Med. Assoc.* 231:1272

10. Tuazon, C. U., Sheagren, J. N. 1974. Increased rate of carriage of *Staphylococcus aureus* among narcotic addicts. *J. Infect. Dis.* 129:725–27
11. Aly, R., Colman, J., Shinefield, H. I., Maibach, H. I. 1978. Comparison of *Staphylococcus aureus* adherence to nasal mucosal cells of infants and adults. *Proc. Int. Congr. Microbiol., 12th,* München, pp. 177, I5
12. Fainstein, V., Musher, D. M. 1980. Bacterial adherence to pharyngeal cells during viral infection. *J. Infect. Dis.* 141:172–76
13. Aly, R., Shinefield, H. R., Litz, C. Maibach, H. I. 1980. Role of teichoic acid in the binding of *Staphylococcus aureus* to nasal epithelial cells. *J. Infect. Dis.* 141:463–65
14. Aly, R., Maibach, H. I., Shinefield, H. R., Mandel, A., Strauss, W. G. 1974. Bacterial interference among strains of *Staphylococcus aureus. J. Infect. Dis.* 129:720–24
15. Shinefield, H. R., Riddle, J. C., Boris, M. 1971. Bacterial interference between strains of *Staphylococcus aureus,* 1960 to 1970. *Am. J. Dis. Child.* 121:148–52
16. Strauss, W. G., Maibach, H. I., Shinefield, H. R. 1969. Bacterial interference: treatment of recurrent furunculosis. *J. Am. Med. Assoc.* 208:861–63
17. Houck, P. W., Nelson, J. D., Kay, J. L. 1972. Fatal septicemia due to *Staphylococcus aureus* 502A. Report of a case and review of the infectious complications of bacterial interference programs. *Am. J. Dis. Child.* 123:45–48
18. Anthony, B. F., Wannamaker, L. W. 1967. Bacterial interference in experimental burns. *J. Exp. Med.* 125:319–36
19. Dajani, A. S., Gray, E. D., Wannamaker, L. W. 1970. Bactericidal substances from *Staphylococcus aureus.* Biological properties. *J. Exp. Med.* 131:1004–15
20. McCabe, W. R. 1968. Role of the reticuloendothelial system in bacterial interference in embryonated eggs. *J. Lab. Clin. Med.* 72:318–28
21. Bröte, L., Niléhn, B. 1976. Wound infections in general surgery with special reference to the occurrence of *Staphylococcus aureus. Scand. J. Infect. Dis.* 8:89–97
22. Meislin, H. W., Lerner, S. A., Graves, M. H., McCehee, M. D., Kocka, F. E., Morrello, J. A., Rosen, P. 1977. Cutaneous abscesses. Anaerobic and aerobic bacteriology and out-patient management. *Ann. Intern. Med.* 87: 145–49
23. Heeb, M. A. 1971. Deep soft tissue abscesses secondary to nonpenetrating trauma. *Surgery* 69:550–53
24. Griffiths, H. E. D., Jones, D. M. 1971. Pyogenic infections of the spine. A review of twenty-eight cases. *J. Bone Joint Surg.* 53:383–91
25. Finland, M., Barnes, M. W. 1978. Bacteremic superinfections of patients with bacteremia: occurrence, bacteriology, mortality, and duration of hospitalization of Boston City Hospital during 12 selected years between 1935 and 1972. *J. Infect. Dis.* 138:829–36
26. Williams, G. T. Houang, E. T., Shaw, E. J., Tabaqchali, S. 1976. Bacteraemia in a London teaching hospital 1966–75. *Lancet* 2:1291–93
27. Pedersen, F. K., Petersen, E. A. 1976. Bacterial endocarditis at Blegdamshospitalet in Copenhagen 1944–73. *Scand. J. Infect. Dis.* 8:99–105.
28. Hers, J. F. P., Goslings, W. R. O., Masurel, N., Mulder, J. 1957. Death from asiatic influenza in the Netherlands. *Lancet* 2:1164–65
29. Martin, C. M., Kunin, C. M., Gottlieb, L. S., Finland, M. 1959. Asian influenza A in Boston. 1957–1958. II. Severe staphylococcal pneumonia complicating influenza. *Arch. Intern. Med.* 103: 532–42
30. Loosli, C. G. 1968. Synergism between respiratory viruses and bacteria. *Yale J. Biol. Med.* 40:522–40
31. Austin, R. M., Daniels, C. A. 1978. The role of protein A in the attachment of staphylococci to influenza infected cells. *Lab. Invest.* 39:128–32
32. Forsgren, A., Sjöquist, J. 1966. "Protein *A*" from *S. aureus.* I. Pseudoimmune reaction with human γ-globulin. *J. Immunol.* 97:822–27
33. Forsgren, A. 1970. Significance of protein A production by staphylococci. *Infect. Immun.* 2:672–73
34. Sjöquist, J. 1973. Structure and immunology of protein A. *Contrib. Microbiol. Immunol.* 1:83–92
35. Forsgren, A. 1969. Protein A from *Staphylococcus aureus.* VIII. Production of protein A by bacterial and L-forms of *Staphylococcus aureus. Acta Pathol. Microbiol. Scand.* 75:481–90
36. Goldstein, E., Lippert, W., Warshaner, D. 1974. Pulmonary alveolar macrophage. Defender against bacterial infection of the lung. *J. Clin. Invest.* 54: 519–28

37. Jakab, G. J., Wan, G. A., Sannes, P. L. 1980. Alveolar macrophage ingestion and phagosome-lysosome fusion defect associated with virus pneumonia. *Infect. Immun.* 27:960–68
38. Larson, H. E., Blades, R. 1976. Impairment of human polymorphonuclear leukocyte function by influenza virus infection. *Lancet* 1:283
39. Kleinerman, E. S., Snyderman, R., Daniels, C. A. 1975. Depressed monocyte chemotaxis during acute influenza infection. *Lancet* 2:1063–66
40. Klebanoff, S. J., Clark, R. A. 1978. *The Neutrophil: Function and Clinical Disorders.* Amsterdam, New York, Oxford: North-Holland. 810 pp.
41. Stossel, T. P. 1975. Phagocytosis: Recognition and ingestion. *Semin. Hematol.* 12:83–116
42. Babior, B. M. 1978. Oxygen-dependent microbial killing by phagocytes. (Two parts). *N. Engl. J. Med.* 298:659–68, 721–25
43. Wilkinson, P. C. 1976. Recognition and response in mononuclear and granular phagocytes. A review. *Clin. Exp. Immunol.* 5:355–66
44. Koenig, M. G. 1972. The phagocytosis of staphylococci. In *The Staphylococci,* ed. J. O. Cohen, 15:365–83. New York: Wiley-Interscience. 548 pp.
45. Humphreys, D. W., Wheat, L. J., White, A. 1974. Staphylococcal heat-stable opsonins. *J. Lab. Clin. Med.* 84: 122–28
46. Verhoef, J., Peterson, P. K., Kim, Y., Sabath, L. D., Quie, P. G. 1977. Opsonic requirements for staphylococcal phagocytosis: Heterogeneity among strains. *Immunology* 33:191–97
47. Wheat, L. J., Humphreys, D. W., White, A. 1974. Opsonization of staphylococci by normal human sera: the role of antibody and heat-labile factors. *J. Lab. Clin. Med.* 83:73–78
48. Verhoef, J., Peterson, P. K., Quie, P. G. 1977. Kinetics of staphylococcal opsonization, attachment, ingestion and killing by human polymorphonuclear leukocytes: a quantitative assay using [^{3}H] thymidine labeled bacteria. *J. Immunol. Methods* 14:303–11
49. Mudd, S., Yoshida, K., Li, I. W., Lenhart, N. A. 1963. Identification of a somatic antigen of *Staphylococcus aureus* critical for phagocytosis by human blood leukocytes. *Nature (London)* 199:1200–1
50. Shayegani, M., Hisatsune, K., Mudd, S. 1970. Cell wall component which affects the ability of serum to promote phagocytosis and killing of *Staphylococcus aureus. Infect. Immun.* 2:750–56
51. Karakawa, W. W., Young, D. A. 1979. Immunological specificity of heat-stable opsonins in immune and nonimmune sera and their interaction with nonencapsulated and encapsulated strains of *Staphylococcus aureus. Infect. Immun.* 25:175–86
52. Verbrugh, H. A., Van Dijk, W. C., Peters, P., Van Erne, M. E., Daha, M. R., Peterson, P. K., Verhoef, J. 1980. Opsonic recognition of staphylococci mediated by cell wall peptidoglycan: antibody-independent activation of human complement and opsonic activity of peptidoglycan antibodies. *J. Immunol.* 124:1167–73
53. Schleifer, K. H., Kandler, O. 1972. Peptidoglycan types of bacterial cell walls and their taxonomic implications. *Bacteriol. Rev.* 36:407–77
54. Seidl, P. H., Schleifer, K. H. 1978. Rapid test for the serological separation of staphylococci from micrococci. *Appl. Environ. Microbiol.* 35:479–82
55. Young, D. A., Dobson, P., Karakawa, W. W. 1979. Immunological specificity of natural opsonins and their role in the cross-reactivity between *Staphylococcus aureus* Mardi and *Escherichia coli* 101. *Infect. Immun.* 25:954–49
56. Stossel, T. P., Field, R. J., Gitlin, J. D., Alper, C. A., Rosen, F. S. 1975. The opsonic fragment of the third component of human complement (C3). *J. Exp. Med.* 141:1329–47
57. Verbrugh, H. A., Van Dijk, W. C., Van Erne, M. E., Peters, R., Peterson, P. K., Verhoef, J. 1979. Quantitation of the third component of human complement attached to the surface of opsonized bacteria: opsonin-deficient sera and phagocytosis-resistant strains. *Infect. Immun.* 26:808–12
58. Verbrugh, H. A., Van Dijk, W. C., Peters, R., Van der Tol, M., Verhoef, J. 1979. The role of *Staphylococcus aureus* cell wall peptidoglycan, teichoic acid and protein A in the processes of complement activation and opsonization. *Immunology* 37:615–21
59. Wilkinson, B. J., Kim, Y., Peterson, P. K., Quie, P. G., Michael, A. F. 1978. Activation of complement by cell surface components of *Staphylococcus aureus. Infect. Immun.* 20:388–92
60. Peterson, P. K., Wilkinson, B. J., Kim, Y., Schmeling, D., Douglas, S. D., Quie, P. G., Verhoef, J. 1978. The key role of peptidoglycan in the opsonization of

Staphylococcus aureus. J. Clin. Invest. 61:597–609

61. Stalenheim, G., Götze, O., Cooper, N. R., Sjöquist, J., Müller-Eberhard, H. J. 1973. Consumption of human complement components by complexes of IgG with protein A of *Staphylococcus aureus. Immunochemistry* 10:501–7
62. Bokish, V. A. 1975. Interaction of peptidoglycans with anti-IgGs and with complement. *Z. Immunitätsforsch. Exp. Klin. Immunol.* 147:320–30.
63. Verbrugh, H. A., Van Dijk, W. C., Peters, R., Van der Tol, M., Peterson, P. K., Verhoef, J. 1979. *Staphylococcus aureus* opsonization mediated via the classical and alternative complement pathways. A kinetic study using MgEGTA chelated serum and human sera deficient in IgG and complement factors C15 and C2 *Immunology* 36: 391–97
64. Ward, P. A., Lepow, I. H., Newman, L. J. 1968. Bacterial factors chemotactic for polymorphonuclear leukoytes *Am. J. Pathol.* 52:725–36
65. Wilkinson, P. C. 1980. Leukocyte locomotion and chemotaxis: effects of bacteria and viruses. *Rev. Infect. Dis.* 2: 293–318
66. Schmeling, D. J., Peterson, P. K., Hammerschmidt, D. E., Kim, Y., Verhoef, J., Wilkinson, B. J., Quie, P. G. 1979. Chemotaxigenesis by cell surface compoments of *Staphylococcus aureus. Infect. Immun.* 26:57–63
67. Griffin, F. M. Jr., Griffin, P. A., Leider, J. E., Silverstein, S. C. 1975. Studies on the mechanism of phagocytosis. I. Requirements for circumferential attachment of particle-bound ligands to specific receptors on the macrophage plasma membrane. *J. Exp. Med.* 142: 1263–82
68. Verhoef, J., Peterson, P. K., Quie, P. G. 1977. Human polymorphonuclear leukocyte receptors for staphylococcal opsonins. *Immunology* 33:231–40
69. Mudd, S. 1970. A successful parasite: infection by *Staphylococcus aureus.* In *Infectious Agents and Host Reactions.* ed. S. Mudd, pp. 197–227. Philadelphia: Saunders
70. Adlam, C., Pearce, J. H., Smith, H. 1970. Virulence mechanisms of staphylococci grown *in vivo* and *in vitro. J. Med. Microbiol.* 3:147–55
71. Koenig, M. G., Melly, M. A., Rogers, D. A. 1962. Factors relating to the virulence of staphylococci. II. Observations on four mouse-pathogenic strains. *J. Exp. Med.* 116:589–99
72. Ekstedt, R. D. 1972. Immunity to the staphylococci. In *The Staphylococci,* ed. J. O. Cohen, 16:385–418. New York: Wiley-Interscience. 548 pp.
73. Laxdal, T., Messner, R. P., Williams, R. C., Quie, P. G. 1968. Opsonic, agglutinating, and complement-fixing antibodies in patients with subacute bacterial endocarditis. *J. Lab. Clin. Med.* 71:638–53
74. Martin, R. R., Daugharty, H., White, A. 1966. Staphylococcal antibodies and hypersensitivity to teichoic acids in man. *Antimicrob. Agents Chemother.* 1965:91–96
75. Taubler, J. H. 1968. Staphylococcal delayed hypersensitivity in mice. I. Induction and *in vivo* demonstration of delayed hypersensitivity. *J. Immunol.* 101:546–49
76. Mudd, S. J., Taubler, H., Baker, A. G. 1970. Delayed hypersensitivity to *Staphylococcus aureus* in human subjects. *J. Reticuloendoth. Soc.* 8:493–98
77. Shayegani, M., De Courcy, S. J., Mudd, S. 1973. Cell-mediated immunity in mice infected with *S. aureus* and elicited with specific bacterial antigens. *J. Reticuloendothel. Soc.* 14:44–51
78. Easmon, C. S. F., Glynn, A. A. 1978. Role of *Staphylococcus aureus* cell wall antigens in the stimulation of delayed hypersensitivity after staphylococcal infection. *Infect. Immun.* 19:341–42
79. Baughn, R. E., Bonventrie, P. F. 1975. Cell-mediated immune phenomena induced by lymphokines from splenic lymphocytes of mice with chronic staphylococcal infection. *Infect. Immun.* 11:313–19
80. Tribble, J. L., Bolen, J. B. 1979. Delayed hypersensitivity to *Staphylococcus aureus* in mice: *in vitro* responses to isolated staphylococcal antigens. *Immunology* 38:819–25
81. Jeljaszewicz, J. 1972. Toxins (Hemolysins). See Ref. 72, pp. 249–80
82. Peterson, P. K., Wilkinson, B. J., Kim, Y., Schmeling, D., Quie, P. G. 1977. Influence of encapsulation on staphylococcal opsonization and phagocytosis by human polymorphonuclear leukocytes. *Infect. Immun.* 19: 943–49
83. Peterson, P. K., Verhoef, J., Sabath, L. D., Quie, P. G. 1977. Effect of protein A on staphylococcal opsonization. *Infect. Immun.* 15:760–64
84. Rotta, J. 1975. Endotoxin-like properties of the peptidoglycan. *Z. Immunitätsforsch. Klin. Exp. Immunol.* 149: 230–44

85. Heymer, B. 1975. Biological properties of the peptidoglycan. *Z. Immunitätsforsch. Klin. Exp. Immunol.* 149: 245–57
86. Hill, M. J. 1968. A staphylococcal aggressin. *J. Med. Microbiol.* 1:33–43
87. Easmon, C. S. F., Hamilton, I., Glynn, A. A. 1973. Mode of action of a staphylococcal anti-inflammatory factor. *Br. J. Exp. Pathol.* 54:638–45
88. Weksler, B. B., Hill, M. J. 1969. Inhibition of leukocyte migration by a staphylococcal factor. *J. Bacteriol.* 98: 1030–35
89. Grov, A. 1976. The effect of specific antibodies on the inhibition of leukocyte migration caused by staphylococcal peptidoglycan. *Acta Pathol. Microbiol. Scand. Sect. B* 84:315–17
90. Lew, P. D., Zubler, R., Vandaux, P., Farquet, J. J., Waldvogel, F. A., Lambert, P. H. 1979. Decreased heat-labile opsonic activity and complement levels associated with evidence of C3 breakdown products in infected pleural effusions. *J. Clin. Invest.* 63:326–34
91. Jacob, H. S., Craddock, P. R., Hammerschmidt, D. E., Moldow, C. F. 1980. Complement-induced granulocyte aggregation. An unsuspected mechanism of disease. *N. Engl. J. Med.* 302:789–94
92. Quie, P. G., Hill, H. R., Davis, A. T. 1974. Defective phagocytosis of staphylococci. *Ann. NY Acad. Sci.* 236:233–42
93. Clark, R. A. 1978. Disorders of granulocyte chemotaxis. In *Leukocyte Chemotaxis: Methods, Physiology and Clinical Complication,* ed. J. I. Gallin, P. G. Quie. New York: Raven.
94. Quie, P. G. 1969. Chronic granulomatous disease of childhood. *Adv. Pediatr.* 16:287–300
95. Goldman, J. M., Catovsky, D. 1972. The function of the phagocytic leucocytes in leukaemia. *Br. J. Haematol.* (Suppl) 23:223–30
96. Schöppfer, K., Douglas, S. D., Wilkinson, B. J. 1980. Immunoglobulin E. antibodies against *Staphylococcus aureus* cell walls in the sera of patients with hyperimmunoglobulinemia E and recurrent staphylococcal infection. *Infect. Immun.* 27:563–68

Ann. Rev. Med. 1981. 32:123–42
Copyright © 1981 by Annual Reviews Inc. All rights reserved

NEUROENDOCRINOLOGICAL EFFECTS OF ALCOHOL

❖7387

Theodore J. Cicero, Ph.D.

Departments of Psychiatry and Anatomy & Neurobiology,
Washington University School of Medicine, Saint Louis, Missouri 63110

INTRODUCTION

Overview

The aim of this review is to discuss the impact of alcohol on the endocrine system and the possible involvement of alcohol-induced endocrine disturbances in the behavioral, pharmacological, and biochemical actions of the drug as well as the biomedical complications associated with its long-term use. Interest in this area stems from the realization of two important factors: alcohol exerts substantial effects on virtually every endocrine system; and because of the pervasive influence of hormones in the cellular biochemistry of every organ, drug-induced alterations in any endocrine system are likely to be significantly involved in the actions of alcohol.

Conceptual and Methodological Issues Involved in Alcohol-Endocrine Studies

PHARMACOLOGICAL AND ENDOCRINOLOGICAL VARIABLES There has been a tendency for many investigators to ignore pharmacological considerations in their investigations of drug-endocrine interactions. For example, many studies examine the endocrine effects of a single dose of drug at one time interval after its injection. This creates misleading results and may in part explain some of the confusion in the literature, since it has been shown that alcohol exerts biphasic effects on many hormones as a function of both dose- and time-response considerations (see below).

Studies in the female have largely been avoided, principally because of the cyclic nature of their hormone secretion rates—the most prominent of which are those associated with the menstrual or estrous cycle, but there

0066-4219/81/0401-0123$01.00

are other episodic hormone secretory patterns as well. It also appears that temporal variation occurs in the secretion of many hormones in the male. Thus, the measured effects of a drug on a specific endocrine system will depend upon when a blood sample is obtained. In addition, there are numerous other variables that might influence endocrine function apart from direct drug effects, including nutritional status, the ingestion of drugs other than alcohol, or the stress associated with the experimental situation.

An additional problem in many studies is that often only a single hormone is measured, and it is assumed that any change in its levels will reflect overall activity in the neuroendocrine system of which it is a part. Since every endocrine system has numerous feedback controls (1), changes in one aspect of the axis often have profound implications for activity in other portions of that axis. Consequently, multiple hormones should be measured in a single blood sample, but this has not always been done.

There are also several important problems encountered in chronic drug studies or when alcoholics are employed to study alcohol's effects on the endocrine system. First, all endocrine systems display remarkable plasticity, and after a period of chronic drug exposure one has no idea whether what one is observing is a direct effect of alcohol on a specific hormone or a secondary alteration in that hormone. Second, chronic alcohol administration frequently results in irreversible organ damage that contaminates any studies of direct drug-induced alterations in endocrine function. Finally, the alcoholic frequently has many problems other than alcoholism, such as poly-drug abuse, and psychiatric and medical difficulties that cloud an interpretation of any endocrine data.

The problems briefly mentioned above should be borne in mind in evaluating the literature on alcohol-endocrine interactions. The degree to which these problems have been recognized has largely dictated the literature selection utilized in the present state-of-the-art review. The interested reader is referred to two recent in-depth reviews in which the literature regarding the variables influencing drug-endocrine studies has been discussed (2, 3).

ALCOHOL-ENDOCRINE INTERACTIONS

Hypothalamic-Pituitary-Gonadal Axis

One general observation can be made about the effects of alcohol on the hypothalamic-pituitary-gonadal axis: it appears to reduce significantly serum testosterone levels after its acute administration in the male of every species studied, and testosterone levels in alcoholics or animals chronically exposed to alcohol are markedly lower than those found in controls (3–10).

There are four mechanisms that could account for reduced serum testosterone levels after acute or chronic alcohol administration. First, alcohol could enhance the metabolism of testosterone by the liver; second, it could block the biosynthesis of the steroid at the level of the testes; third, it could suppress the synthesis and/or release of luteinizing hormone (LH) from the pituitary and thereby inhibit LH-dependent testicular steroidogenesis; or, finally, it could inhibit the release of LH-releasing hormone (LH-RH) from the hypothalamus, resulting in a series of secondary effects on the pituitary and testes. These possibilities will be discussed in the ensuing sections.

LIVER It is well documented that the clearance of testosterone is markedly enhanced relative to controls in the chronic alcoholic human or animal (6–9, 11, 12). This effect appears to be due to an induction of the enzymes responsible for the degradation of testosterone in the liver, primarily 5-α-reductases (6, 11, 12). However, recent studies by Gordon and his colleagues (13, 14) add some complexity to this issue. These investigators found that prolonged administration of alcohol ($>$ 1 year) to baboons or man resulted in a decrease in the activity of 5-α-reductases (13, 14). They suggested that this may be why the metabolism of testosterone is often depressed in humans subjected to very prolonged alcohol consumption (15, 16), and they speculated that a reduced conversion of testosterone to dihydrotestosterone in the secondary sex organs, resulting from decreased 5-α-reductase activity (13, 14), could further exacerbate the effects of alcohol on testosterone-dependent systems.

Additional alterations in the metabolic fate of testosterone after prolonged administration of alcohol in humans and animals have also been demonstrated. For example, chronic alcohol administration has been associated with increases in circulating estrogens (17–19), and this increase in plasma estrogen levels correlates with an increase in hepatic aromatase activity (18), an enzyme involved in converting androgens to estrogens. Moreover, there is an increased rate of conversion of testosterone and androstenedione to their respective estrogens in vivo in man (14, 15, 17). Since estrogen administration promotes signs of "feminization," it seems likely that increases in the formation of estrogens, coupled with decreases in serum testosterone levels, may play a significant role in the disturbances in reproductive endocrinology in the chronic alcoholic.

TESTES A good deal of evidence suggests that alcohol exerts direct effects on the production of testosterone by the testes in the male of every species after both acute and chronic alcohol administration. As mentioned above, both acute and chronic alcohol administration lower serum testosterone levels in the male of all species thus far examined, and, with repeated and persistent usage, alcohol may ultimately irreversibly damage the ultrastructural and biochemical architecture of the testes (see references above; 8, 20–22).

Several in vivo and in vitro studies have demonstrated unequivocally that alcohol directly inhibits LH-dependent testicular steroidogenesis. For example, several groups of investigators have shown that alcohol markedly suppressed gonadotropin-stimulated testosterone production in the male rodent in vivo (18, 23, 24). Since gonadotropin levels were far above those required to promote testicular steroidogenesis in these studies, it seems clear that alcohol exerts direct effects at the level of the testes—whether or not it has any effect on the hypothalamic-pituitary-LH axis (see below).

In several in vitro preparations, alcohol inhibited the basal or gonodotropin-stimulated production of testosterone by the testes in normal male rats (24–29). Moreover, Gordon et al (18) showed that testes obtained from rodents chronically treated with alcohol responded very poorly to gonadotropin stimulation. It should be noted, however, that extremely high concentrations of alcohol have generally been required in vitro to produce significant inhibitions of testicular steroidogenesis, whereas very low doses of alcohol have been required to inhibit steroidogenesis in vivo (18, 23, 24). Although some of this difference might reflect inherent differences in the sensitivity of in vitro versus in vivo preparations, the results of a number of recent studies suggest that the metabolism of alcohol, which occurs readily in vivo and minimally in vitro (25), may be responsible for the effects of alcohol on steroidogenesis. Two consequences of the metabolism of alcohol could be involved: a change in the $NAD^+/NADH$ ratio (since some of the enzymes involved in testicular steroidogenesis are NAD^+-dependent) or the generation of the highly reactive metabolite acetaldehyde. Sufficient evidence has now accumulated to implicate both mechanisms in the actions of alcohol on testicular steroidogenesis (3, 24, 25, 29–32). Whether both are equally involved or whether one predominates under in vivo conditions remains to be determined.

The step in the biosynthetic pathway of testosterone which is inhibited by alcohol or acetaldehyde has been examined by two groups of investigators (31, 32). Gordon et al (31) provided evidence that alcohol inhibits the enzyme converting pregnenolone to progesterone in vivo (Δ^5–3β–dehydrogenase). The mechanism underlying this effect appears to be a depletion of NAD^+ due to the metabolism of alcohol. On the other hand, under in vitro conditions, Cicero et al (32) found that alcohol and acetaldehyde significantly affected only a single step in the biosynthesis of testosterone—the conversion of androstenedione to testosterone. None of the other precursors of testosterone was affected by the two drugs, including the conversion of pregnenolone to progesterone [the step implicated by Gordon et al (31)]. Based upon these studies, it appears that alcohol and acetaldehyde directly affect the biosynthetic pathway for testosterone in the testes by two mechanisms. Both drugs directly affect the conversion of androstenedione to testosterone (in vitro), and this effect may be compounded in vivo (but not in vitro) by a reduction in the conversion of pregnenolone to progesterone.

Hypothalamic-pituitary-LH axis Although the studies described in the preceding section would seem to indicate that the inhibitory effects of alcohol on serum testosterone levels can be attributed exclusively to an action at the level of the testes, it is equally clear from a variety of studies that alcohol has a significant effect on the hypothalamic-pituitary-LH axis.

Several investigators have demonstrated in the rat that acute injections of alcohol significantly depressed serum LH levels (2–5, 33, 34). Chronic alcohol administration also depresses serum LH levels in the rodent (2, 3, 20, 34). The situation in the human regarding the acute and chronic effects of alcohol on serum LH is not as clear as it is in the rat. Increases, decreases, and frequently no changes in serum LH levels have been found after acute or chronic administration of alcohol (2, 3, 6, 15, 21, 35–38). Although an initial interpretation of these data could be that alcohol exerts no impact on the hypothalamic-pituitary-LH axis, at least in the human, these results in fact argue for a strong effect on the central aspect of the hypothalamic-pituitary-gonadal axis. That is, it is well established that the hypothalamic-pituitary-LH axis is continuously inhibited by testosterone, and when the axis is relieved of this negative feedback control, serum LH levels rise by 10- to 20-fold with respect to controls (1). Since in all of the studies in the human discussed above, serum LH levels were only modestly changed, even though serum testosterone levels were markedly reduced, it appears that the hypothalamic-pituitary-LH axis must be significantly inhibited by alcohol. To directly assess this strong inference, several investigators examined the effects of alcohol on the castration-induced rise in serum LH levels in the male rat (5, 33, 34, 39). In these studies alcohol markedly suppressed the increase in serum LH following castration, demonstrating unequivocally that alcohol has very substantial effects upon the hypothalamic-pituitary-LH axis.

The bulk of the currently available evidence suggests that, within the hypothalamic-pituitary axis, alcohol exerts its effects by acting on the hypothalamus. For example, acute or chronic alcohol administration does not reduce the rise in serum LH produced by systemically administered LH-RH in rats and humans (5, 15, 20, 23, 33). Similarly, the effects of LH-RH on the release of LH by the anterior pituitary in vitro are not attenuated in the presence of relatively high concentrations of alcohol (5, 23). Furthermore, if one blocks the rise in serum LH induced by castration with alcohol and then administers LH-RH, LH levels return promptly to the pre-alcohol-treated levels (5, 23, 33, 39). These data demonstrate that LH-RH can overcome alcohol's blockade of the hypothalamic-pituitary-LH axis; hence the drug must be exerting its effects at a suprasellar site within this axis. However, Van Thiel and his associates (40) reported that the response to LH-RH provocation in chronic alcoholics was impaired in some cases; Baker et al (15) also found that a small percentage of alcoholics responded poorly to LH-RH stimulation. It is not clear why these investigators found a pituitary effect of alcohol in some alcoholics, whereas other investigators have not (see above), but perhaps the chronicity of exposure and/or other intervening variables (e.g. nutrition or alterations in other endocrine systems) may explain these discordant findings. Nevertheless, it appears that alcohol can directly alter hypothalamic function (a view not seriously disputed by any investigator), apart from whether or not it has any effects on the pituitary. Most likely alcohol affects the hypothalamus by inhibiting LH-RH release, but this has not yet been unequivocally established.

From the foregoing review of the effects of alcohol on the liver, testes, and hypothalamic-pituitary-LH axis it is apparent that the drug exerts multiple effects on the hypothalamic-pituitary-gonadal axis, resulting in a marked and pronounced impairment of reproductive endocrinology and physiology in the male. It is not known to what extent alcohol also influences the function of the hypothalamic-pituitary-gonadal axis in the female, since there are relatively few available studies which would shed any light on this issue. What is known is that alcohol appears to disrupt the menstrual cycle in the human and estrous cycle in the rodent (e.g. 2, 3) and appears to be a significant gonadal (ovarian) toxin (41), but there have been too few systematic studies undertaken to draw any meaningful conclusions.

CONSEQUENCES OF DRUG-INDUCED ALTERATIONS IN THE HYPOTHALAMIC-PITUITARY-GONADAL AXIS The consequences of alterations in the hormones secreted by the hypothalamic-pituitary-gonadal axis as a function of acute or chronic alcohol administration have not been systematically examined. Recently, however, alcohol-induced reductions in serum testosterone levels were implicated in the development of metabolic tolerance to alcohol. Israel and his colleagues (42, 43), and subsequently Cicero et al (44), reported that castration resulted in markedly enhanced levels of alcohol dehydrogenase in the liver and a corresponding increase in the clearance of alcohol in the male rat. This effect occurred very soon after castration and essentially reached a maximum within a matter of two to three days (44). Moreover, daily injections of testosterone readily reversed this increase in alcohol dehydrogenase and the clearance of alcohol. Since alcohol markedly suppresses serum testosterone levels, acutely and chronically (see above), both groups of investigators speculated that chronic alcohol administration might be equivalent to "chemical" castration. Some support for this hypothesis has recently been provided by Israel et al (42). They reported that metabolic tolerance to alcohol could not be produced in castrated animals, whereas a significant increase in alcohol dehydrogenase and the metabolism of alcohol occurred in sham-operated rats after a period of chronic alcohol administration. These data were interpreted as supporting the notion that drug-induced reductions in testosterone were responsible for the production of metabolic tolerance to alcohol—at least in the male rat.

The possible involvement of testosterone in the control of alcohol dehydrogenase leads to several interesting predictions regarding cross-tolerance between alcohol and other sedative-hypnotic drugs. Many drugs that are cross-tolerant with alcohol also depress serum testosterone levels in the male of several species (e.g. 2, 3, 45). Cicero et al (44) examined whether chronic treatment with morphine increased alcohol dehydrogenase and the clearance of alcohol. They found that morphine produced a pronounced (85%) increase in liver alcohol dehydrogenase activity and a greatly enhanced clearance of alcohol, which was completely reversed by daily injections of testosterone. These results support the hypothesis that liver alcohol dehydrogenase is under the control of testosterone and, furthermore, suggest that the testosterone-depleting effects of many abused substances may be an important biochemical mechanism mediating cross-tolerance between them. It should be clear, however, that all of the studies reviewed above used only the male rat; it remains to be determined whether similar results will be obtained in other species, including man.

With respect to the consequences of changes in the secretion of LH or LH-RH induced by alcohol, there are no data available at the present time.

Hypothalamic-Pituitary-Adrenal Axis

DIRECT EFFECTS OF ALCOHOL ON THE HYPOTHALAMIC-PITUITARY-ADRENAL AXIS Acute alcohol administration at appropriate doses elevates corticosterone levels in animals and cortisol levels in normal human volunteers (46–52), but this has not been observed in all studies (3, 53, 54). The reasons for the latter discrepancies are not entirely clear, but they may be related to dose-response problems, which will be dealt with in depth below. Nevertheless, it appears that in general after very high, intoxicating doses of alcohol in both animals and humans, serum corticosterone and cortisol, respectively, are elevated. With respect to chronic effects of alcohol on the hypothalamic-pituitary-adrenal axis, many groups have shown persistent elevations in cortisol (in man) during the drinking phase and corticosterone (in rats) after acute challenges with alcohol once a critical blood level is achieved (50, 53–62). Moreover, although tolerance develops to the effects of alcohol on the hypothalamic-pituitary-adrenal axis in some studies (60–62), most investigators have failed to find this (3, 48, 54, 55, 57).

With respect to basal, resting levels of cortisol in the alcoholic, increases have generally been found in those subjects still drinking (see references above), whereas levels appear to be normal in the abstinent alcoholic (54, 57, 59). The effects of acute challenges with alcohol in the abstinent alcoholic have also been examined, but conflicting results have been obtained. For example, Merry & Marks (50) administered alcohol to acutely withdrawn alcoholics and found a *depression* in plasma cortisol levels, rather than the increase usually observed after acute administration in normals or alcoholics (see above). On the other hand, two groups observed that resumption of alcohol use in abstinent alcoholics produced increases in cortisol levels comparable to those found in normal subjects or drinking alcoholics (57–59). At present it is unclear what accounts for these differences in results obtained with abstinent alcoholics.

Alcohol does not appear to act directly at the level of the adrenal to enhance the secretion of steroids, but rather the locus of action seems to be on the hypothalamic-pituitary axis. This conclusion is based on the following observations: (*a*) pituitary ACTH levels were depleted after acute alcohol treatment in rodents (61), which appeared to correlate with increased corticosterone release; (*b*) alcohol was ineffective in promoting the release of corticosterone in hypophysectomized animals, which indicates that an intact pituitary is necessary for an effect to be observed (48, 49); and (*c*) in humans with pituitary tumors in which there is clinical evidence of impaired ACTH release, alcohol does not raise cortisol levels (52).

A cautionary note should be made with regard to the effects of alcohol or any drug on the hypothalamic-pituitary-adrenal axis. That is, it is extremely difficult to state in any study of drug-induced changes in this axis whether one is observing a direct effect of the drug per se or a secondary consequence of the stress associated with acute or chronic drug treatment. In the case of alcohol, unfortunately, a strong case can be made that its effects upon the hypothalamic-pituitary-adrenal axis are due exclusively to the stress associated with its administration. This conclusion is based on the observation that in most studies in humans or animals a release of cortisol or corticosterone seems to have occurred in an all-or-none fashion once an intoxicating blood level of alcohol was achieved, and there was no apparent correlation between blood alcohol levels and corticosterone release (e.g. 49, 52–54, 57–59). In addition, the route by which alcohol is administered and how fast the peak blood alcohol levels are attained determine whether an adrenal response occurs at all following acute alcohol administration (63, 64). If alcohol exerts a drug-specific effect on this axis, a good dose-response relationship should exist; it is very difficult to imagine that the rate of onset of peak blood alcohol levels or the route of administration would be important variables if stress were not involved in some way. Until these issues are resolved, it is very difficult to conclude anything with respect to the effects of alcohol per se on the hypothalamic-pituitary-adrenal axis.

CONSEQUENCES OF DRUG-INDUCED ALTERATIONS IN HYPOTHALAMIC-PITUITARY-ADRENAL ACTIVATION The consequences of acute alterations in corticosterone or cortisol production induced by acute alcohol administration in animals or humans, respectively, are not at all well understood. Indeed, there are essentially no data available regarding this point.

With respect to the development of tolerance to and physical dependence on alcohol, however, there does appear to be some participation of the hypothalamic-pituitary-adrenal axis. Sze et al (65, 66) demonstrated in rats and mice that adrenalectomy markedly decreased the severity of the alcohol withdrawal reaction. In other studies Kakihana (46) showed that mice with very high serum corticosterone levels had much more severe withdrawal seizures than did mice with low corticosterone levels. This study seems to corroborate the results of Sze et al (65, 66) and suggests that the basal level of activity in the hypothalamic-pituitary-adrenal axis determines in part the chronic effects of ethanol on the central nervous system.

Several groups of investigators have also found significant interactions between the hypothalamic-pituitary-adrenal axis and the development of tolerance. For example, Sze (67) reported that adrenalectomy abolished the increase in liver alcohol dehydrogenase occurring as a result of chronic alcohol administration in mice. He suggested, therefore, that enhanced levels of activity in the hypothalamic-pituitary-adrenal axis played a "permissive role" in the development of metabolic tolerance to alcohol. In addition, two groups of workers have examined whether alcohol-induced alterations in the hypothalamic-pituitary-adrenal axis participate in the development of "functional" tolerance to alcohol. Tabakoff & Yanai (68) found that treatment of rats with cortolexone, a corticosterone antagonist, resulted in an attenuation of the development of tolerance to the hypothermic effects of alcohol following chronic maintenance on alcohol-containing liquid diets. Similarly, Wood (69) found that chronic treatment with dexamethasone, a corticosterone-like compound, accelerated the development of tolerance to the behavioral effects of alcohol, in addition to antagonizing its acute effects to some extent. These three reports thus seem to provide good evidence that alterations in hypothalamic-pituitary-adrenal activity may play some role in the acute actions of alcohol and the development of metabolic and functional tolerance.

Hypothalamic-Pituitary-Thyroid Axis

DIRECT EFFECTS OF ALCOHOL ON THE HYPOTHALAMIC-PITUITARY-THYROID AXIS The most consistent effect of alcohol on the function of the thyroid is the modest decrease of serum thyroxine (T_4) levels and the marked decrease of serum triiodothyronine (T_3) levels (70–75). This effect has been observed after acute alcohol administration but is most pronounced in chronic alcoholics, particularly those with alcoholic hepatitis or cirrhosis (70–75). It appears, however, that the very low concentrations of T_3 in the serum of alcoholics, and perhaps normals after acute alcohol administration, does not reflect a direct action on the thyroid, since T_4 levels and serum TSH levels are only modestly (T_4) or not at all (TSH) lower than those found in controls (70–75). The most likely mechanism involved appears to be a direct effect of alcohol on the peripheral conversion of T_4 to T_3 (70), which takes place primarily in the liver (76).

Additional studies of the effects of acute or chronic alcohol administration on thyroid hormones have revealed that there is an increased uptake of T_3 and T_4 into a number of target organs, particularly the liver, both in vivo and in vitro (70, 72–74). Moreover, alcohol has also been reported to increase the uptake of iodine by the thyroid (77), but, unfortunately, consistent results have not been obtained (65, 73, 77).

Alcohol does not appear to exert any important effects on the hypothalamic-pituitary aspects of the hypothalamic-pituitary-thyroid axis: when serum TSH levels were measured, the levels of this pituitary hormone were generally unaltered after acute or chronic alcohol administration (see above; 70, 75).

Although the foregoing studies suggest that alcoholics have relatively normal thyroid function, Goldberg (78, 79) reported that chronic alcoholics had a high degree of hypothyroidism, and he suggested that this was a significant feature of their clinical status. Moreover, he claimed a good clinical response to treatment with thyroid hormones in alcoholics (78, 79). On the basis of the large amount of data gathered since these initial reports, however, it does not appear that alcoholics have hypothyroidism or that thyroid hormones have any utility in the treatment of alcoholism (see above; 80–82).

CONSEQUENCES OF ALTERED THYROID FUNCTION Recent studies show that the administration of TRH to drug-naive male rats significantly antagonized the acute effects of alcohol (for a review, see 83). Apart from these studies, however, literature does not indicate that alcohol-induced alterations in the function of the thyroid participate in the drug's acute effects or in the development of tolerance and withdrawal behavior.

However, Israel and his colleagues recently provided evidence of an extremely important interaction between the thyroid and alcohol-induced liver disease (42, 73, 84–86). Several years ago these investigators showed that chronic alcohol consumption produced a hypermetabolic condition in the liver which was accompanied by an increase in the rate of oxygen consumption (73, 84, 86). Since thyroid hormones also produce a hypermetabolic state (70, 73, 84), they suggested that alterations in thyroid hormones may play a role in alcoholic hepatitis and cirrhosis. In support of this hypothesis, they found that thyroidectomy virtually eliminated the hypermetabolic state induced by chronic alcohol administration in male rats (84, 86) and that the administration of the antithyroid drug propylthiouracil (PTU) reversed the effects of alcohol on the liver (85, 86).

On the basis of these studies in animals, Orrego et al (71) and Israel et al (42, 70) conducted double-blind clinical tests of the efficacy of PTU as a treatment modality in patients with alcoholic liver disease. In agreement with much of the work described above, they found that T_3 levels were markedly reduced in their alcoholic population (70) when compared to controls, and there was an excellent inverse correlation between serum T_3 levels and the extent of liver damage (70, 71). The administration of PTU was found to provide an excellent degree of reversal of liver damage in chronic alcoholics and was most effective in patients with the lowest serum T_3 levels and most severe liver damage (70, 71). Patients with normal T_3 levels did not respond effectively to PTU treatment. These studies are extremely important for two reasons. First, serum T_3 levels may provide a good screening device to assess alcoholic liver damage, although it should be noted that other conditions also give rise to reduced serum T_3 (70), which may limit its usefulness; second, PTU may provide the first clinically efficacious treatment for one of the most devastating consequences of alcoholism.

Growth Hormone

DIRECT EFFECTS OF ALCOHOL ON GROWTH HORMONE RELEASE

The evidence regarding the effects of acute and chronic alcohol administration on basal growth hormone levels in humans is equivocal at the present time. Various investigators have found increases, decreases, or no changes in steady-state growth hormone levels in human males following acute alcohol administration (3, 87–90). But, there appears to be good agreement that alcohol acutely blocks stimulated growth hormone release (e.g. by environmental stimuli, hypoglycemia, or L-dopa chemotherapy) in normal males and chronic alcoholics (87–93). In animals there are essentially no systematic studies of the effects of alcohol on growth hormone release. There have been no attempts to localize the actions of alcohol on growth hormone release to the hypothalamus or pituitary.

CONSEQUENCES OF DRUG-INDUCED ALTERATIONS IN GROWTH HORMONE None of the studies in the literature indicate that changes in growth hormone secretion after acute or chronic alcohol administration are involved in the acute or chronic actions of alcohol. It seems obvious, however, that persistent alcohol-induced depressions in the secretion of this critical anterior pituitary hormone would have significant effects upon many biochemical processes in virtually every organ in the body.

Prolactin

DIRECT EFFECTS OF ALCOHOL ON PROLACTIN RELEASE There are few studies of the effects of alcohol on prolactin release in the male human. Most investigators have found increases in prolactin levels in normal volunteers after acute alcohol administration and in the drinking chronic alcoholic, but it should be noted that not all investigators have observed such changes (3, 19, 38, 94–98). Furthermore, resting levels of prolactin in the abstinent chronic alcoholic have been found to be normal or slightly reduced in the few studies available (99, 100).

In terms of the effects of alcohol on stimulated prolactin release, Ylikahri et al (38, 101) found that the prolactin response to TRH (increase) was significantly enhanced during periods of peak blood alcohol levels in normal volunteers. On the other hand, two groups of investigators found that TRH-stimulated prolactin release was blunted during withdrawal from alcohol in chronic alcoholics (100) or 10 hours after the administration of a very high dose of alcohol ("the hangover period") in normal human volunteers (38, 101).

Taken as a whole, these acute and chronic studies in the human seem to indicate that (*a*) acute alcohol administration tends to increase prolactin in normal volunteers; (*b*) basal, resting levels in the alcoholic are normal or modestly lower than those found in controls; and (*c*) the prolactin response to TRH is enhanced by acute alcohol administration but may be attenuated during the "hangover period" or in the chronic alcoholic undergoing withdrawal. It should be obvious, however, that these conclusions are general, tentative ones, and not all of the data are consistent with these interpretations (see above). In the rat chronic alcohol administration appears to elevate serum prolactin levels (3), but this issue has not been examined in any depth. There are no studies available in which acute alcohol administration in the rodent has been examined.

In terms of the locus of action of alcohol within the hypothalamic-pituitary axis, there are essentially no data but it has been reported that alcohol increases the release of prolactin by the anterior pituitary in vitro (102).

A major problem in interpreting the significance of drug-induced alterations in prolactin is that it is difficult to differentiate drug-specific effects from the nonspecific effects of stress. Since prolactin levels rise dramatically in response to stress in many species (103) and alcohol itself may be a stressor, particularly when given in intoxicating amounts, the significance of alcohol-induced changes in prolactin must be interpreted with caution.

CONSEQUENCES OF DRUG-INDUCED ALTERATIONS IN PROLACTIN Changes in prolactin levels subsequent to acute or chronic alcohol administration are not well documented, so it is not surprising that there has been little speculation about their significance. Van Thiel et al (19, 95) suggested that increases in prolactin levels in chronic alcoholics might participate in the "feminization," particularly gynecomastia, observed in these individuals. Williams (94) suggested that alcohol-induced increases in prolactin may contribute to a higher degree of breast cancer in alcoholics. At present, however, there are no convincing data that unequivocally establish this relationship.

Vasopressin

DIRECT EFFECTS OF ALCOHOL ON VASOPRESSIN RELEASE There seems to be little question that acute alcohol administration or ingestion produces diuresis in both the male and female animal and human (104–109). It seems equally clear that alcohol produces this effect by inhibiting the secretion of vasopressin. This conclusion is based on two sets of observations. First, alcohol-induced diuresis can be overcome by the administration of synthetic vasopressin (104, 106, 109); second, serum vasopressin levels are markedly reduced in normal human volunteers and animals after acute alcohol administration on the ascending slope of the blood alcohol curve when diuresis is most prominent (105, 107, 110–113). During chronic alcohol administration little or no tolerance seems to develop to the antidiuretic effects of alcohol or its inhibition of vasopressin release (107, 113). The actions of alcohol on vasopressin release have also been studied during alcohol withdrawal in chronic alcoholics or on the descending limb of the blood alcohol curve after acute administration. Generally, "rebound" increases in vasopressin have been found in these studies (105, 107, 110).

The locus of alcohol's actions on vasopressin release appears to be at or above the supraoptic nucleus of the hypothalamus. The evidence in support of this conclusion can be summarized as follows. First, stimuli that normally give rise to vasopressin release, such as a sodium chloride load (21, 104) or direct application of acetylcholine to neurons in the supraoptic nucleus in the hypothalamus (105), can override the diuretic effects of alcohol. Second, alcohol inhibits the histological changes in vasopressin-containing cells of the supraoptic nucleus produced by a large dose of sodium chloride in animals (114). Finally, alcohol inhibits the electrically evoked discharges of the supraoptic secretory cells in the hypothalamus (115). This action is not due to a direct effect of alcohol on these cells, however, since they respond very well to direct stimulation (see the first and second points above). Alcohol's effects on vasopressin release therefore seem to be due to an action at some neuronal locus above the neurosecretory cells.

CONSEQUENCES OF ALCOHOL-INDUCED ALTERATIONS IN VASOPRESSIN LEVELS Although there are few data linking changes in vasopressin to the acute pharmacological effects of alcohol (aside from diuresis), there is a growing body of evidence suggesting that vasopressin and other centrally active peptides may play a significant role in the development of tolerance and physical dependence. Hoffman and co-workers (116, 117) reasoned that vasopressin might participate in the acquisition and maintenance of tolerance to alcohol because of two observations. First, many peptides in brain, including vasopressin, have significant biological activity, particularly with respect to the acquisition and retention of certain learned behaviors in rats. Second, vasopressin and its analogs alter the acute effects of psychoactive drugs and the development of tolerance to narcotics (see 116–118 for reviews).

In view of these observations and the fact that alcohol alters serum vasopressin levels (see above), Hoffman et al (116, 117) examined whether the continuous administration of arginine vasopressin (AVP) during a period of chronic alcohol exposure would alter either the acquisition or maintenance of tolerance to alcohol. They found that tolerance to the hypothermic effects of alcohol persisted for longer periods of time following discontinuation of alcohol treatment in AVP-treated rats than it did in control-injected animals. Moreover, they demonstrated that only the retention, and not the acquisition, of tolerance was affected by AVP administration (i.e. AVP-treated animals attained the same level of tolerance at apparently the same rate). The observations of Hoffman et al (116, 117) have subsequently been replicated by Rigter & Crabbe (119, 120).

The effects of vasopressin on alcohol withdrawal symptoms have also been examined. On this point, however, there is some controversy. Hoffman et al (116, 117) reported that vasopressin did not affect withdrawal signs and reactions in their animals. Rigter & Crabbe (119, 120) found that the intraventricular injection of vasopressin markedly increased both tolerance persistence and withdrawal symptomatology. There is no apparent explanation for this difference.

On the basis of the studies of Hoffman et al and Rigter & Crabbe it appears that alcohol-induced alterations in vasopressin may participate in the development of tolerance and physical dependence. These are potentially important observations, and further studies are needed to more fully characterize the mechanisms underlying this interaction.

CONCLUSIONS

In this review an attempt has been made to describe the available studies concerning the impact of alcohol on a variety of neuroendocrine systems. As should be apparent, we know relatively little about the effects of alcohol on the full range of neuroendocrine systems. A good deal is known about the effects of alcohol on the hypothalamic-pituitary-gonadal axis and much less so on all of the remaining neuroendocrine axes. Clearly, we need to know a great deal more about the effects of alcohol on the hypothalamic-pituitary-adrenal and hypothalamic-pituitary-thyroid axes and upon growth hormone and vasopressin release, as well as several other neuroendocrines that have been virtually ignored. Moreover, a number of questions remain with respect to the effects of alcohol on all of the neuroendocrine systems discussed in this review.

The significance of alcohol-induced changes in endocrine state are only now being appreciated. This is somewhat surprising in view of the fact that hormones play such an integral role in the cellular biochemistry and physiology of every organ in the body. It is virtually unimaginable that significant alcohol-induced alterations in the secretion of any hormone, therefore, would not be profoundly significant for both the acute pharmacological actions of the drug and the long-term consequences associated with its use (i.e. biomedical complications and the development of tolerance and physical dependence). Indeed, investigators exploring the ramifications of alterations in neuroendocrine function as a by-product of acute and chronic alcohol administration find very significant alcohol-endocrine interactions (see appropriate preceding sections). This strongly validates the assumption that changes in endocrine state are integrally involved in alcohol's acute and chronic effects in many organ systems. Future studies along these lines should be quite informative and important.

Literature Cited

1. Fink, G. 1979. Feedback actions of target hormones on hypothalamus and pituitary with special reference to gonadal steroids. *Ann. Rev. Physiol.* 41: 571–85
2. Cicero, T. J. 1980. Common mechanisms underlying the effects of ethanol and narcotics on neuroendocrine function. In *Advances in Substance Abuse,* ed. N. K. Mello, pp. 201–54. Greenwich, Conn: JAI
3. Cicero, T. J. 1980. Sex differences in the effects of psychoactive drugs on endocrine function: clinical and experimental evidence. In *Research Advances in Alcohol and Drug Problems,* ed. Y. Israel, O. Kalant, H. Kalant. New York: Plenum. In press
4. Cicero, T. J., Badger, T. M. 1977. Effects of alcohol on the hypothalamic-pituitary-gonadal axis in the male rat. *J. Pharmacol. Exp. Ther.* 201:427–33
5. Cicero, T. J., Bernstein, D., Badger, T. M. 1978. Effects of acute alcohol administration on reproductive endocrinology in the male rat. *Alcohol Clin. Exp. Res.* 2:249–54
6. Gordon, G. G., Southren, A. L., Altman, K., Rubin, E., Lieber, C. S. 1976. The effect of alcohol (ethanol) administration on sex hormone metabolism in normal men. *N. Engl. J. Med.* 295: 793–97
7. Gordon, G. G., Southren, A. L., Lieber, C. S. 1978. The effects of alcoholic liver disease and alcohol ingestion on sex hormone levels. *Alcohol. Clin. Exp Res.* 2:259–64
8. Lester, R., Van Thiel, D. H. 1977. Gonadal function in chronic alcoholic men. *Adv. Exp. Med. Biol.* 85A:399–414
9. Southren, A. L., Gordon, G. G. 1976. Effects of alcohol and alcoholic cirrhosis on sex hormone metabolism. *Fertil. Steril.* 27:202
10. Ylikahri, R., Huttunen, M. 1974. Hangover and testosterone. *Br. Med. J.* 2:445
11. Rubin, E., Lieber, C. S., Altman, K., Gordon, G. G., Southren, A. L. 1976. Prolonged ethanol consumption increases testosterone metabolism in the liver. *Science* 191:563–64
12. Bode, C., Martini, G. A., Bode, J. C. 1978. Effect of alcohol on microsomal cortisol 4 en-5α-reductase in the liver of rats fed on a standard or low protein diet. *Horm. Metab. Res.* 10:63–64
13. Gordon, G. G., Vittek, J., Ho, R., Rosenthal, W. S., Southren, A. L., Lieber, C. S. 1980. The effect of chronic alcohol use on hepatic 5α-A-ring reductase in the baboon and man. *Gastroenterology.* In press
14. Gordon, G. G., Southren, A. L., Lieber, C. S. 1979. Hypogonadism and feminization in the male: a triple effect of alcohol. *Alcohol Clin. Exp. Res.* 3:210–12
15. Baker, H., Berger, H. G., Dekretser, D. M., Dulmanis, A., Hartson, B., O'Connor, S., Paulsen, C. A., Purecell, N., Rennie, G. C., Seah, C. S., Taft, H. P., Wang, C. 1976. A study of the endocrine manifestations of hepatic cirrhosis. *Q. J. Med.* 45:145–78
16. Southren, A. L., Gordon, G. G., Olivo, J., Rafii, F., Rosenthal, W. S. 1973. Androgen metabolism in cirrhosis of the liver. *Metabolism* 22:695–702
17. Gordon, G. G., Olivo, J., Rafii, F., Southren, A. L. 1975. Conversion of androgens to estrogens in cirrhosis of the liver. *J. Clin. Endocrinol. Metab.* 40: 1018–26
18. Gordon, G. G., Southren, A. L., Vittek, J., Lieber, C. S. 1979. The effect of alcohol ingestion on hepatic aromatase activity and plasma steroid hormone in the rat. *Metabolism* 28:20–24
19. Van Thiel, D. H., Gavaler, J. S., Lester, R., Loriaux, D. L., Braunstein, G. D. 1975. Plasma estrone, prolactin, neurophysin and sex steroid-binding globulin in chronic alcoholic men. *Metabolism* 24:1015–19
20. Symons, A. M., Marks, V. 1975. Effects of alcohol on weight gain and the hypothalamic-pituitary-gonadotropin axis in the maturing male rat. *Biochem. Pharmacol.* 24:955–58
21. Van Thiel, D. H., Lester, R. 1976. Alcoholism: its effect on hypothalamic-pituitary-gonadal function. *Gastroenterology* 71:318–26
22. Van Thiel, D. H., Gavaler, J. S., Lester, R., Goodman, M. D. 1975. Alcohol-induced testicular atrophy. An experimental model for hypogonadism occurring in chronic alcoholic men. *Gastroenterology* 69:326–32
23. Cicero, T. J., Meyer, E. R., Bell, R. D. 1978. Effects of ethanol on the hypothalamic-pituitary-luteinizing hormone axis and testicular steroidogenesis. *J. Pharmacol. Exp. Ther.* 208:210–15
24. Ellingboe, J., Varanelli C. C. 1979. Ethanol inhibits testosterone biosynthesis by direct action on leydig cells. *Res. Commun. Chem. Pathol. Pharmacol.* 24:87–102

25. Cicero, T. J., Bell, R. D., Meyer, E. R., Badger, T. M. 1980. Ethanol and acetaldehyde directly inhibit testicular steroidogenesis. *J. Pharmacol. Exp. Ther.* 213:228–33
26. Badr, S. M., Bartke, A., Dalterio, S., Bulger, W. 1977. Suppression of testosterone production by ethyl alcohol. Possible mode of action. *Steroids* 30:647–55
27. Cobb, C. F., Ennis, M. F., Van Thiel, D. H., Gavaler, J. S., Lester, R. 1980. Isolated testes perfusion: a method using a cell- and protein-free perfusate useful for the evaluation of potential drug and/or metabolic injury. *Metabolism* 29:71–79
28. Cobb, C. F., Ennis, M. F., Van Thiel, D. H., Gavaler, J. S., Lester, R. 1979. Alcohol: its effect on the isolated perfused rat testes. *Alcohol Clin. Exp. Res.* 3:170
29. Cobb, C. F., Van Thiel, D. H., Ennis, M. F., Gavaler, J. S., Lester, R. 1978. Acetaldehyde and ethanol are testicular toxins. *Gastroenterology* 75:958
30. Deleted in proof
31. Gordon, G. G., Vittek, J., Southren, A. L., Munnangi, P., Lieber, C. S. 1980. Effect of chronic alcohol ingestion on the biosynthesis of steroids in rat testicular homogenate *in vitro*. *Endocrinology* 106:1880–85
32. Cicero, T. J., Bell, R. D. 1980. Effects of ethanol and acetaldehyde on the biosynthesis of testosterone in the rodent testes. *Biochem. Biophys. Res. Commun.* 94:814–19
33. Chapin, R. E., Breese, G. R., Mueller, R. D. 1980. Possible mechanisms of reduction of plasma luteinizing hormone by ethanol. *J. Pharmacol. Exp. Ther.* 212:6–10
34. Cicero, T. J., Badger, T. M. 1977. A comparative analysis of the effects of narcotics, alcohol and the barbiturates on the hypothalamic-pituitary-gonadal axis. *Adv. Exp. Med. Biol.* 85:95–115
35. Mendelson, J. H., Mello, N. K., Ellingboe, J. 1977. Effects of acute alcohol intake on the pituitary-gondadal hormone in normal human males. *J. Pharmacol. Exp. Ther.* 202:676–82
36. Mendelson, J. H., Ellingboe, J., Mello, N. K., Kuehnle, J. 1978. Effects of alcohol on plasma testosterone and luteinizing hormone levels. *Alcohol. Clin. Exp. Res.* 2:255–58
37. Van Thiel, D. H., Lester, R. 1978. Further evidence for hypothalamic-pituitary dysfunction in alcoholic men. *Alcohol. Clin. Exp. Res.* 2:265–70
38. Ylikahri, R. H., Huttunen, M. O., Harkonen, O., Leino, T., Helenius. T., Liewendahl, K., Karonen, S. L. 1978. Acute effects of alcohol on anterior pituitary secretion of the tropic hormones. *J. Clin. Endocrinol. Metab.* 46:715–20
39. Redmond, G. P. 1980. Effect of ethanol on pulsatile gonadotropin secretion in the male rat. *Alcohol. Clin. Exp. Res.* 4:226
40. Van Thiel, D. H., Lester, R., Sherins, R. J. 1974. Hypogonadism in alcoholic liver disease: evidence for a double defect. *Gastroenterology* 67:1188–99
41. Van Thiel, D. H., Gavaler, J. S. Lester, R. 1977. Ethanol: a gonadal toxin in the female. *Drug Alcohol Depend.* 2:373
42. Israel, Y., Khanna, J. M., Orrego, H., Rachamin, G., Wahid, S., Britton, R., Macdonald, A., Kalant, H. 1979. Studies on metabolic tolerance to alcohol, hepatomegaly and alcoholic liver disease. *Drug Alcohol Depend.* 4:109–29
43. Rachamin, G., Macdonald, A., Wahid, S., Clapp, J. J., Khanna, J. M., Israel, Y. 1980. Modulation of alcohol dehydrogenase and ethanol metabolism by sex hormones in the spontaneously hypertensive rat. *Biochem. J.* 186:483–90
44. Cicero, T. J., Bernard, J. D., Newman, K. 1980. Effects of castration and chronic morphine administration on liver alcohol dehydrogenase and the metabolism of ethanol in the male rat. *J. Pharmacol. Exp. Ther.* In press
45. Cicero, T. J. 1980. Effects of exogenous and endogenous opiates on the hypothalamic-pituitary-gonadal axis in the male. *Fed. Proc.* 39:2551–54
46. Kakihana, R. 1977. Endocrine and autonomic studies in mice selectively bred for different sensitivity to ethanol. *Adv. Exp. Med. Biol.* 85A:83–95
47. Suzuki, T., Higashi, R., Hiros, T., Iheda, H., Tamura, H. 1972. Adrenal 17-hydroxycorticosteroid secretion in the dog in response to ethanol. *Acta Endocrinol.* 70:736
48. Ellis, F. W. 1966. Effect of ethanol on plasma corticosterone levels. *J. Pharmacol. Exp. Ther.* 153:121–27
49. Czaja, C., Kalant, H. 1961. The effect of acute alcoholic intoxication on adrenal ascorbic acid and cholesterol in the rat. *Can. J. Biochem. Physiol.* 39:327–34
50. Merry, J., Marks, V. 1969. Plasma hydrocortisone response to ethanol in chronic alcoholics. *Lancet* 1:921–23
51. Bellet, S., Roman, L., DeCastro, O. A. P., Herrera, M. 1970. Effects of acute ethanol intake on plasma 11-hydroxycorticosteroid levels. *Metabolism* 19:644

52. Jenkins, J. S., Connolly, J. 1968. Adrenocortical response to ethanol in man. *Br. Med. J.* 11:804–5
53. Kissin, B., Schenker, V., Schenker, A. C. 1960. The acute effect of alcohol ingestion on plasma and urinary 17-hydroxycorticoids in alcoholic subjects. *Am. J. Med. Sci.* 239:690–705
54. Mendelson, J. H., Stein, S. 1966. Serum cortisol levels in alcoholic and nonalcoholic subjects during experimentally induced ethanol intoxication. *Psychosom. Med.* 28:616–26
55. Tabakoff, B., Jaffe, R., Ritzmann, R. 1978. Corticosterone concentrations in mice during ethanol drinking and withdrawal. *J. Pharm. Pharmacol.* 30: 371–74
56. Kakihana, R., Moore, J. 1975. Circadian rhythm of corticosterone in mice: the effect of chronic consumption of alcohol. *Psychopharmacology* 46: 301–5
57. Stokes, P. E. 1973. Adrenocortical activation in alcoholics during chronic drinking. *Ann. NY Acad. Sci.* 215:77
58. Merry, J., Marks, V. 1972. The effect of alcohol, barbiturate, and diazepam on hypothalamic/pituitary/adrenal function in chronic alcoholics. *Lancet* 2: 990–91
59. Mendelson, J. H., Ogata, M., Mello, N. K. 1971. Adrenal function and alcoholism. I. Serum cortisol. *Psychosom. Med.* 33:145–57
60. Crossland, J., Ratcliffe, F. 1968. Some effects of chronic alcohol administration in the rat. *Br. J. Pharmacol.* 32: 413–14
61. Noble, E. P., Kakihana, R., Butte, J. C. 1971. Corticosterone metabolism in alcohol-adapted mice. In *Biological Aspects of Alcohol,* ed. M. K. Roach, W. M. McIsaac, P. J. Creaven, pp. 389–417. Austin: Univ. Tex. Press
62. Kakihana, R., Butte, J. C., Hathaway, A., Noble, E. P. 1971. Adrenocortical response to ethanol in mice: modification by chronic ethanol consumption. *Acta Endocrinol.* 67:653–54
63. Kalant, H., Hawkins, R. D., Czaja, C. 1963. Effect of acute alcohol intoxication on steroid output of rat adrenals *in vitro. Am. J. Physiol.* 204:849–55
64. Kalant, H. 1975. Direct effects of ethanol on the nervous system. *Fed. Proc.* 34:1930–41
65. Sze, P. Y., Yanai, J., Ginsburg, B. E. 1974. Adrenal glucocorticoids as a required factor in the development of ethanol withdrawal seizures in mice. *Brain Res.* 80:155
66. Sze, P. 1977. The permissive role of glucocorticoids in the development of ethanol dependence and tolerance. *Drug Alcohol Depend.* 2:381–96
67. Sze, P. Y. 1975. The permissive role of glucocorticoids in the induction of liver alcohol dehydrogenase by ethanol. *Biochem. Med.* 14:156–61
68. Tabakoff, B., Yanai, J. 1979. Cortexolone antagonizes development of alcohol tolerance in mice. *Psychopharmacology* 64:123–24
69. Wood, W. 1977. Facilitation by dexamethasone of tolerance to ethanol in the rat. *Psychopharmacology* 52:67–72
70. Israel, Y., Walfish, P. G., Orrego, H., Blake, S., Kalant, H. 1979. Thyroid hormones in alcoholic liver disease. *Gastroenterology* 76:116–22
71. Orrego, H., Kalant, H., Israel, Y., Blake, J., Medline, A., Rankin, J. G. Armstrong, A., Kapur, B. 1979. Effect of short-term therapy with propylthiouracil in patients with alcoholic liver disease. *Gastroenterology* 76:105–15
72. Ramakrishana, S., Prasanna, C. V., Balasubramanian, A. 1976. Effect of alcohol intake on rat hepatic enzymes and thyroid function. *Indian J. Biochem. Biophys.* 13:49–51
73. Israel, Y., Videla, L., Macdonald, A., Bernstein, J. 1973. Metabolic alterations produced in the liver by chronic ethanol administration. Comparison between the effects produced by ethanol and by thyroid hormones. *Biochem. J.* 134:523–29
74. Bleecker, M., Ford, D. H., Rhines, R. K. 1969. A comparison of ^{131}I-triiodothyronine accumulation and degradation in ethanol-treated and control rats. *Life Sci.* 8:267
75. Green, J. R. B., Snitcher, E. J., Mowat, N. A. G., Ekins, R. P., Rees, L. H., Dawson, A. M. 1977. Thyroid function and thyroid regulation in euthyroid men with chronic liver disease: evidence of multiple abnormalities. *Clin. Endocrinol.* 7:453–61
76. Oppenheimer, J. H., Bernstein, G., Surks, M. I. 1968. Increased thyroxine turnover and thyroidal function after stimulation of hepatocellular binding of thyroxine. *J. Clin. Invest.* 47:1399–1406
77. Murdock, H. R. 1967. Thyroidal effect of alcohol. *Q. J. Stud. Alcohol* 28: 419–23
78. Goldberg, M. 1960. The occurrence and treatment of hypothyroidism among alcoholics. *J. Clin. Endocrinol. Metab.* 20:609–21

79. Goldberg, M. 1962. Thyroid function in chronic alcoholism. *Lancet* 2:746–49
80. Selzer, M. L., Van Houten, N. H. 1964. Normal thyroid function in chronic alcoholism. *J. Clin. Endocrinol. Metab.* 24:380–82
81. Kalant, H., Sereny, G., Charlebois, R. 1962. Evaluation of triiodothyronine in the treatment of acute alcohol intoxication. *N. Engl. J. Med.* 267:1–14
82. Satterfield, J. H., Guze, S. B. 1961. Treatment of alcoholic patients with triiodothyronine. *Dis. Nerv. Syst.* 22: 227
83. Nemeroff, C., Loosen, P., Bissette, G., Manberg, P., Wilson, I., Lipton, M., Prange, A. 1979. Pharmaco-behavioral effects of hypothalamic peptides in animals and man: focus on thyrotropin release hormone and neurotensin. *Psychoneuroendocrinology* 3:279–310
84. Israel, Y., Videla, L., Bernstein, J. 1975. Liver hypermetabolic state after chronic ethanol consumption: hormonal interrelationships and pathogenic implications. *Fed. Proc.* 34: 2052–59
85. Israel, Y., Kalant, H., Orrego, H., Khanna, J. M., Videla, L., Phillips, J. M. 1975. Experimental alcohol-induced hepatic necrosis: suppression by propylthiouracil. *Proc. Natl. Acad. Sci. USA* 72:1137–41
86. Bernstein, J., Videla, L., Israel, Y. 1975. Hormonal influences in the development of the hypermetabolic state of the liver produced by chronic administration of ethanol. *J. Pharmacol. Exp. Ther.* 192:583–91
87. Bagdade, J. D., Gale, C. C., Porte, D. 1972. Hormone-fuel interrelationships during alcohol hypoglycaemia in man. *Proc. Soc. Exp. Biol. Med.* 141:540–42
88. Arky, R., Freinkel, N. 1964. The response of plasma growth hormone to insulin and ethanol induced hypoglycaemia in two patients with "isolated adrenocorticotrophic defect." *Metabolism* 13:547
89. Andreani, D., Tamburrano, G., Javicoli, M. 1976. Alcohol hypoglycaemia: hormonal changes. In *Hypoglycaemia: Proc. Eur. Symp. Horm. Metab. Res. Suppl. Ser.,* ed. D. Andreani, P. Lefebure, V. Marks, pp. 99–105, Stuttgart: Thieme
90. Bellet, S., Yoshimine, N., Decastro, O. A. P., Roman, L., Parmar, S. S., Sandberg, H. 1970. Effect of alcohol ingestion on growth hormone levels: their relation to 11-hydroxycorticoid levels and serum FFA. *Metabolism* 20:762
91. Ganda, D. P., Savin, C. T., Iber, F., Glennon, J. A., Mitchell, M. L. 1978. Transient suppression of growth hormone secretion after chronic ethanol intake. *Alcohol. Clin. Exp. Res.* 2:297–99
92. Priem, H. A., Shanley, B. C., Malan, C. 1976. Effect of alcohol administration on plasma growth hormone response to insulin-induced hypoglycemia. *Metabolism* 25:397–403
93. Riesco, J., Costamillere, L., Litvak, J. 1974. Growth hormone secretion in chronic alcoholics: lack of response to L-dopa and glucagon. *Rev. Med. Chile* 102:443
94. Williams, R. R. 1976. Breast and thyroid cancer and malignant melanoma promoted by alcohol-induced pituitary secretion of prolactin, TSH and MSH. *Lancet* 1:996–99
95. Van Thiel, D. H., McClain, C. J., Elson, M. K., McMillin, M. J. 1978. Hyperprolactinemia and thyrotropin-releasing factor (TRH) responses in men with alcoholic liver disease. *Alcohol. Clin. Exp. Res.* 2:344–48
96. Earll, J. M., Gaunt, K., Earll, L., Djuh, Y. Y. 1976. Effect of ethyl alcohol on ionic calcium and prolactin in man. *Aviat. Space Environ. Med.* 47:808–810
97. Cushman, P. Jr., Kreek, M. J. 1974. Methadone-maintained patients. Effect of methadone on plasma testosterone, FSH, LH, and prolactin. *NY State J. Med.* 74:1970–73
98. Turkington, R. W. 1972. Serum prolactin levels in patients with gynecomastia. *J. Clin. Endocrinol. Metab.* 34:62
99. Wright, J. 1978. Endocrine effects of alcohol. *Clin. Endocrinol. Metab.* 7: 351–67
100. Loosen, P. T., Prange, A. J. Jr. 1977. Alcohol and anterior-pituitary secretion (letter). *Lancet* 2:985
101. Ylikahri, R. H., Huttenen, M. O., Harkonen, M. 1976. Effect of alcohol on anterior pituitary secretion of trophic hormones. *Lancet* 1:1353
102. Thorner, M. O., Kirk, C. R., MacLeod, R. M. 1978. Alcohol stimulation of prolactin release from perfused isolated rat pituitary cells. *Fed. Proc.* 37:637
103. Carroll, B. J. 1978. Neuroendocrine function in psychiatric disorders. In *Psychopharmacology: A Generation of Progress,* ed. M. A. Lipton, A. DiMascio, K. F. Killam, pp. 487–97. New York: Raven
104. Kleeman, C. R. 1972. Water metabolism. In *Clinical Disorders of Fluid and Electrolyte Metabolism,* ed. H. Max-

well, C. R. Kleeman, pp. 243–57. New York: McGraw-Hill

105. Bisset, G. W., Walker, J. M. 1957. The effect of nicotine, hexamethonium and ethanol on the secretion of antidiuretics and oxytocic hormones of the rat. *Br. J. Pharmacol.* 12:461–67
106. Eggleton, M. G. 1942. Diuretic action of alcohol in man. *J. Physiol.* 101: 172–91
107. Marquis, C., Marchetti, J., Burlet, C., Boulange, M. 1975. Secretion urinaire et hormone antidiuretique chez des rats soumis a une administration repetee d'ethanol. *CR Seances Soc. Biol.* 169: 154–61
108. Cobo, E., Quintero, C. A. 1969. Milk-ejecting and anti-diuretic activities under neurohypophyseal inhibition with alcohol and water overload. *Am. J. Obstet. Gynecol.* 105:877–87
109. Haggard, H. W., Greenberg, L. A., Carroll, P. P. 1941. Studies on the absorption, distribution and elimination of alcohol. *J. Pharmacol. Exp. Ther.* 71: 349–57
110. Linkola, J., Ylikahri, R., Fyhrquist, F., Wallenius, M. 1978. Plasma vasopressin in ethanol intoxication and hangover. *Acta Physiol. Scand.* 104:180–87
111. Linkola, J., Fyhrquist, F., Forsander, O. 1977. Effects of ethanol on urinary arginine vasopressin excretion in two rat strains selected for their different ethanol preferences. *Acta Physiol. Scand.* 101:126–28
112. Helderman, J., Vestal, R., Rowe, J., Tobin, J., Andres, R., Robertson, G. 1978. The response of arginine vasopressin to intravenous ethanol and hypertonic saline in man: the impact of aging. *J. Gerontol.* 33:39–47
113. Sereny, G., Rapaport, A., Hudson, H. 1966. The effect of alcohol withdrawal on electrolyte and acid-based balance. *Metabolism* 15:896
114. Raiha, N. 1960. Effect of ethanol on cytological changes induced by salt load in nucleus supraopticus of rat. *Proc. Soc. Exp. Med. Biol.* 103:387–91
115. Miller, D. A., Mill, J. A. Jr. 1967. Interactions among ethanol, hypothermia and asphyxia in guinea pigs. *Cryobiology* 3:400
116. Hoffman, P., Ritzmann, R., Walter, R., Tabakoff, B. 1978. Arginine vasopressin maintains ethanol tolerance. *Nature* 276:614–16
117. Hoffman, P., Ritzmann, R., Tabakoff, B. 1979. The influence of arginine vasopressin and oxytocin on ethanol dependence and tolerance. In *Currents in Alcoholism,* 5:5–16. New York: Grune & Stratton
118. Rigter, H., Crabbe, J. 1979. Modulation of memory by pituitary hormones and related peptides. *Vitam. Horm.* 37:153–241
119. Rigter, H., Crabbe, J. 1980. Neurohypophyseal peptides and ethanol. In *Hormones and the Brain.* Geneva: Int. Health Found. In press
120. Crabbe, J. C., Rigter, H. 1980. Hormones, peptides and ethanol responses. In *Alcohol Tolerance and Dependence,* ed. H. Rigter, J. C. Crabbe, pp. 219–313, Amsterdam: Elsevier/North-Holland

Ann. Rev. Med. 1981. 32:143–60
Copyright © 1981 by Annual Reviews Inc. All rights reserved

ANEMIA OF PREMATURITY

❖7388

Peter R. Dallman, M.D.

Department of Pediatrics, M-650, University of California,
San Francisco, California 94143

INTRODUCTION

During the first two months after birth, the concentration of hemoglobin declines rapidly from the highest to the lowest values that occur at any period of development (1–6). The low point is termed the physiologic anemia of infancy. It is considered physiologic because it is characteristic of healthy infants, it occurs in the absence of any recognized nutritional deficiency, and there is no sustained response to blood transfusion. If the fall in hemoglobin concentration were an isolated phenomenon, it would significantly diminish the availability of oxygen to the tissues. However, while hemoglobin concentration is decreasing, there are other changes that lead to a progressive rightward shift of the oxygen equilibrium curve, permitting more oxygen to be released by hemoglobin at any given oxygen tension (7–9). While these modifications in oxygen delivery are taking place, there are concurrent changes in oxygen requirements. As the newborn adapts to an extrauterine environment, his oxygen needs progressively increase (10–12), and he becomes less able to withstand oxygen lack (13). Therefore, if tissue oxygenation is to remain adequate, oxygen extraction and/or cardiac output must increase.

In the otherwise healthy term infant, the adaptations in oxygen delivery during the first two months after birth are rarely associated with clinical evidence of hypoxia. However, in the premature infant, the fall in hemoglobin concentration is more profound, and in some infants clinical evidence suggests that O_2 delivery is compromised and that a blood transfusion is required (14–16). Nevertheless, the rapidity of developmental changes that affect oxygen delivery and the complexity of their interactions has been an impediment to developing clear-cut criteria for transfusion. Consequently, this is an area in which clinical practices vary widely. Regardless

0066-4219/81/0401-0143$01.00

of the type of management, the anemia of prematurity is a self-limited condition since changes in both hemoglobin concentration (5) and oxygen affinity (7) favor enhanced oxygen delivery after 2–3 months of age. However, hypoxia in some infants during the nadir of their anemia might have both acute and long-lasting consequences.

This paper first reviews the normal postnatal changes in the concentration of hemoglobin and in oxygen delivery, and outlines some of the mechanisms believed to be responsible for these changes. Next, there is the difficult matter of deciding under what conditions the anemia becomes pathologic and should be reversed by transfusion. Last to be considered are factors that can accentuate the anemia of prematurity and that can be either avoided or minimized.

Changes in Concentration of Hemoglobin

The postnatal decline in concentration of hemoglobin can be attributed largely to an abrupt increase in oxygen availability. At birth, the concentration of hemoglobin in umbilical cord blood averages about 17 g/dl (17). This relative polycythemia results from the hypoxic intrauterine environment, in which blood in the most highly oxygenated fetal vessel, the umbilical vein, is far less saturated with oxygen than blood that has passed through the lungs after birth (18). Fetal red cell production is elevated (19) presumably in response to an increased activity of erythropoietin (20, 21), the humoral stimulator of red cell production, which occurs in response to tissue hypoxia. The high rate of intrauterine red cell production is reflected at birth by average values of about 5% reticulocytes and 7 nucleated red cells/per 100 white cells in term infants. Values in premature infants are often 2–3 times higher (17).

After birth, as oxygen from the lungs virtually saturates the hemoglobin in arterial blood, oxygen delivery to the tissues is facilitated. This results in diminished activity of erythropoietin in the plasma (20) and a decrease in the production of red cells during the first few days of life (2, 22). In the peripheral blood, these changes are reflected in a precipitous fall in the reticulocyte count and the disappearance of nucleated red blood cells (2, 17). In the bone marrow, the percentage of erythroid precursor cells declines from an average of 35% at birth to 15% after one week (23, 24), and extramedullary hematopoiesis comes to a halt.

A marked depression in the rate of red blood cell production continues for 4–6 weeks (2, 24). During this period, the rate of red cell destruction is greater than in later life: red cell life span is estimated to be less than 70 days compared with 120 days in the adult (19, 25). As a consequence of decreased red cell production, more rapid destruction, and a marked increase in blood volume with growth, the mean hemoglobin concentration in term infants drops to the lowest value of 11.0 g/dl at about 2 months of age (26). The fall in hemoglobin concentration is of greater magnitude in the low birth weight infant (5). At about 2 months, infants with birth weights between 1500 and 2000 g have a mean hemoglobin concentration of about 9.5 g/dl; the corresponding value for those with a birth weight below 1500 g is 9.0 g/dl (Figure 1). This decline in hemoglobin concentration is independent of iron nutrition and occurs at a time when neonatal iron stores are still present (2, 5).

After about 4–6 weeks of age, increased red blood cell production is indicated by a rise in the reticulocyte count (2), and hemoglobin concentration in the term infant rises gradually to maintain a mean hemoglobin value

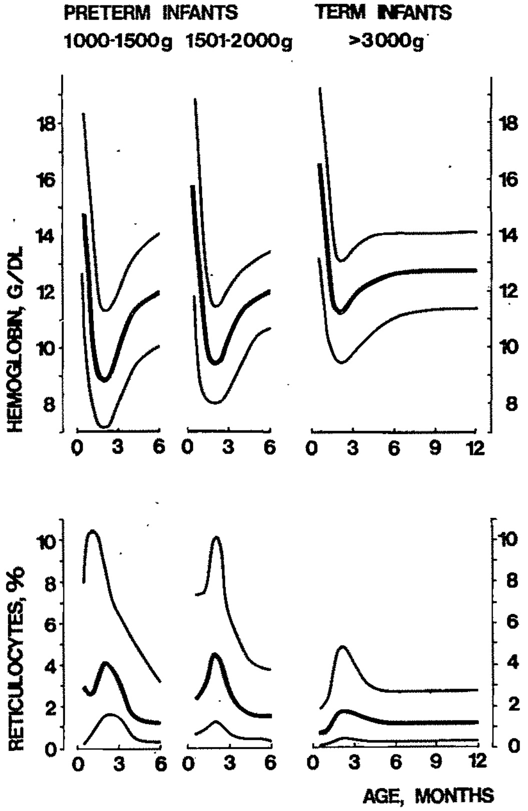


Figure 1 Hemoglobin concentration and reticulocyte counts in preterm and term infants. Median values and 95% confidence limits are indicated for each of three birth weight categories. (Hemoglobin values are from Lündstrom et al 1977 (5) and Saarinen & Siimes 1978 (26); reticulocyte values are from unpublished data on the same infants from whom hemoglobin values were derived.)

of 12.5 g/dl through the remainder of the first year of life (26). The maintenance of this hemoglobin concentration requires a doubling of body iron by one year of age (27). This additional iron must be supplied from dietary sources after 4–6 months of age when neonatal iron stores have been depleted (2, 28).

In the premature infant, the rise in reticulocyte count (2) and the rate of increase in hemoglobin concentration after 2 months of age are even greater than in the term infant, and by 5 months the mean values of the premature are equivalent to those of term infants (29). The hemoglobin values in the premature infant become dependent on an adequate supply of dietary iron as early as 2 months after birth (1–3, 5), because neonatal iron stores are proportionally smaller and the relative rate of growth is greater than in the term infant (a three- to fourfold increase in weight in the 4–5 months after birth during which the term infant doubles his birth weight).

Changes in the Oxygen Affinity of Hemoglobin

The hemoglobin molecule combines reversibly with oxygen, which allows red blood cells to pick up oxygen from the lungs and then deliver it to the tissues. The oxygen affinity of hemoglobin assumes particular importance in the interpretation of anemia in infants because it undergoes major changes after birth, and there are large individual variations at any given age (7, 8). Oxygen affinity is most readily depicted in quantitative terms by the oxygen dissociation curve, which is derived by plotting the oxygen saturation of a red cell suspension against its oxygen tension or partial pressure (Po_2) under standard in vitro conditions of pH and temperature. The oxygen tension at which hemoglobin becomes 50% saturated is termed the P_{50}. In adults, the P_{50} averages 27 torr (mm Hg). The oxygen dissociation curve of the fetus and newborn is shifted to the left (an average P_{50} of 18–22 torr), with the lowest values in the most premature infants (30); this denotes a higher oxygen affinity that favors the extraction of oxygen from the maternal circulation. Although the high oxygen affinity of fetal blood suits the conditions of the intrauterine environment, it is not as favorable for the release of oxygen to the tissues at the increased oxygen tensions that prevail after birth. The adaptation to an oxygen dissociation curve more favorable for extrauterine conditions is not reached for many weeks. By 3 months of age, the curve has shifted to the right, to a position similar to that of the adult; here it remains relatively constant throughout the remainder of life (7, 8). The experimental use of adult blood for exchange transfusion of hypoxic newborn infants (31) is an attempt to facilitate the release of oxygen to the tissues by speeding a change that would normally occur more slowly. The procedure, though theoretically appealing, remains controversial in terms of clinical effectiveness; it is not discussed in this review since it is applicable primarily to the immediate postnatal period, before the anemia of prematurity has developed.

The basis for the gradual shift of the oxygen dissociation curve after birth was not understood until about 15 years ago when it was found that certain organic phosphates can decrease the affinity of hemoglobin A for oxygen by competing for oxygen-binding sites (32–34). The most important of these compounds in modulating the interaction between hemoglobin and oxygen is 2, 3-diphosphoglycerate (DPG). Because hemoglobin F has a lower affinity for DPG than hemoglobin A, it binds oxygen more tenaciously (7–9, 35, 36), which accounts for an oxygen dissociation curve in the newborn that lies to the left. The postnatal rise in P_{50} with a shift of the oxygen dissociation curve to the right is partly attributable to the gradual replacement of hemoglobin F by hemoglobin A and a consequent increase in the capacity of hemoglobin to bind DPG (7, 8). The switch from production of hemoglobin F to hemoglobin A starts *in utero* and is largely complete by 4 months after birth, but the rate of transition is slow and is little influenced by the timing of birth (37). Infants who have been transfused are likely to have a low oxygen affinity for their age because the percentage of hemoglobin F has been decreased by the hemoglobin A in the transfused blood (6, 38).

There is also a second mechanism by which DPG can influence the oxygen affinity of hemoglobin. The DPG molecule is acidic. Thus a high concentration of DPG decreases the pH within the red blood cell, which then shifts the oxygen dissociation curve to the right. This change in oxygen affinity with an alteration in pH is called the Bohr effect (39). At baseline concentrations of red cell DPG, binding of DPG to the hemoglobin A molecule appears to be the more important influence on oxygen affinity; at high levels of DPG the pH effect predominates (9, 40). Usually, DPG and hemoglobin are present in approximately equimolar concentrations in the red blood cell (34); however, following the experimental initiation of a hypoxic stimulus, the concentration of DPG increases over a period of about 48 hours (41). This response is brought about by a variety of factors that alter the rates of synthesis and utilization of DPG (for review, see 40). Premature infants of low birth weight have elevated mean red cell DPG values between 6 and 10 weeks (7), when the anemia of prematurity is most evident. Consequently, the decrease in oxygen affinity that facilitates oxygen delivery during this period may be largely mediated by the effect of DPG on red cell pH.

There are other factors that influence red cell pH and modify the affinity of hemoglobin for oxygen but that are not evident from in vitro measurements of oxygen affinity. Probably the most important of these is the almost instantaneous and reversible change in oxygen affinity that occurs each time the red cell makes a circuit through the vasculature (34). As blood becomes progressively more acid with the release of carbon dioxide and metabolic acids into the capillaries, the affinity of hemoglobin for oxygen decreases through the Bohr effect, facilitating the release of oxygen to the tissues. This mechanism favors varying degrees of oxygen release in different parts of the circulation, since metabolically active and hypoxic tissues will provide the most acidic environments for hemoglobin. Such local regulation of oxygen release is of great physiologic importance but is difficult to evaluate clinically, since pH is subject to rapid changes in different parts of the circulation and will depend on variations in regional blood flow and rates of tissue metabolism.

Oxygen Delivery to Tissues

Studies of oxygen delivery in infants have dealt primarily with the interaction between the concentration of hemoglobin and its in vitro affinity for oxygen. This interaction is depicted in Figure 2 where the oxygen-carrying capacity of the blood (which is proportional to the concentration of hemo-

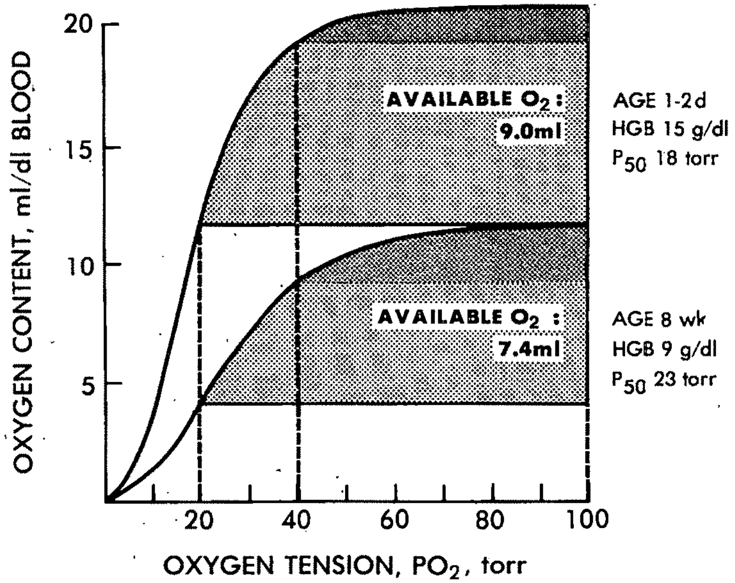


Figure 2 Oxygen equilibrium curves of blood from a representative premature infant, 1–2 days and 8 weeks after birth. The curves are based on mean values for infants with birth weights between 1000 and 1500 g obtained by Delivoria-Papadopoulos et al., 1971 (7). The *Oxygen-unloading capacity,* indicated in dark gray, is greater at 8 weeks than at 1–2 days after birth. *Available oxygen* is represented by both the dark and light gray areas; values are lower at 8 weeks (42), thus apparently corresponding more closely to clinical circumstances.

globin) is plotted against oxygen tension. The volume of oxygen unloaded to the tissues can be estimated by determining the decrease in oxygen-carrying capacity as oxygen tension decreases from values characteristic of arterial blood to the values present in the venous side of the circulation (7, 8, 42).

In the newborn, blood has a high oxygen-carrying capacity, but the ability to release oxygen to the tissues is impaired by a high oxygen affinity. Between 1 and 3 months of age, the oxygen-carrying capacity of the blood is decreased, but this apparent handicap is partially compensated by the enhanced capacity of hemoglobin to release oxygen to the tissues (7, 8, 42). Thus, the ability to deliver oxygen during the postnatal period can remain relatively constant despite major changes in concentration of hemoglobin and oxygen affinity.

If anemia is defined exclusively in reference to hemoglobin values in healthy subjects of the same age and maturity, then we should consider the low hemoglobin concentration in most premature infants between 1 and 3 months of age to be normal and the term anemia to be inappropriate. However, it may be more reasonable to define anemia in terms of physiologic function, in reference to the adequacy of the oxygen supply to the tissues (7, 14, 42, 43). Efforts have been made to define anemia on a functional basis, taking both the concentration of hemoglobin and the position of the oxygen dissociation curve into account. Even though such definitions may be difficult to apply to individual patients, they are conceptually useful because they avoid the widespread tendency to consider oxygen delivery only in terms of the concentration of hemoglobin.

One term proposed is the *oxygen-releasing capacity* of blood, calculated from the difference in oxygen content between arterial blood at 95% oxygen saturation and mixed venous blood at an oxygen tension of 40 torr (7). In healthy premature infants the oxygen releasing capacity actually increases during the first 2 months after birth because the decrease in oxygen affinity of the blood more than counterbalances the decline in hemoglobin concentration (Figure 2) (7). Unfortunately, the maintenance of a stable or increasing oxygen-releasing capacity may obscure the fact that *reserves* for oxygen delivery have diminished and that there may be impaired ability to further compensate for the increased demands of respiratory exertion, low environmental temperature, and other factors that could increase oxygen requirements.

Another defect of the oxygen-releasing capacity is the assumption that mixed venous Po_2 remains constant. However, data obtained from newborn lambs, which have a postnatal anemia similar to that in infants, show that the mixed venous Po_2 does not remain constant; it decreases in an adaptive manner to as low as 20 torr, to maintain a stable oxygen consumption (44). The same conclusion was reached in experiments on rhesus monkeys in which an exchange transfusion was performed with stored autologous blood that had become DPG-depleted and had a high oxygen affinity (45). No significant change in oxygen consumption or cardiac output was observed; the major adaptation for the maintenance of oxygen delivery was a substantial decrease in mixed venous P_{O_2}.

These considerations favor the use of another expression of oxygen delivery, the *available oxygen* (14, 42). Available oxygen is calculated from the difference in oxygen content between an arterial P_{O_2} of 100 torr and a mixed venous P_{O_2} of 20 torr; it represents the amount of oxygen that *can* be extracted before the need arises for an increase in cardiac output at rest. The definition of available oxygen is based on evidence in animals and in man that when oxygen tension falls to about 20 torr cardiac output increases in response to oxygen demands (46, 47). An oxygen tension of at least 20 torr is also believed to be necessary for a sufficient oxygen gradient to exist between the capillary and the interior of tissue cells (48). Available oxygen can be calculated to decrease slightly between the first days after birth and about 8 weeks of age (Figure 2); this decrease is in accord with clinical evidence that, at least in some infants, there is little reserve for maintenance of oxygen delivery between 6 and 12 weeks of age (14-16). Available oxygen in a group of infants believed to have clinical signs of anemia was significantly lower than in those without such findings, even though the difference in hemoglobin concentration between the two groups did not reach statistical significance (14). Unfortunately, individual values for available oxygen showed marked overlap between infants with signs of anemia and those without; this makes it impractical to rely on this calculation as the sole criterion in deciding whether an individual infant requires a blood transfusion. However, it may prove to be a good and practical way to express the infant's ability to deliver oxygen. Recently developed instruments can provide an oxygen dissociation curve on a few microliters of capillary blood. The results would make it a relatively simple matter to calculate an acutal *available oxygen* rather than relying on assumed values (42).

Redistribution of Blood Flow

In response to hypoxia, there is a redistribution of blood flow within the body while total cardiac output and arterial blood pressure tend to remain constant. These circulatory adjustments to hypoxia are generally beneficial in that they augment the flow of blood to organs with a great dependence on oxygen for their metabolism. The mechanisms that mediate the redistribution of blood flow are discussed in a recent review (49) and fall primarily into two categories: local vascular effects and effects of chemoreceptor stimulation.

The *local vascular effects* have been studied by perfusion experiments in experimental animals. Local perfusion with hypoxic blood resulted in marked vasodilation in the heart and the brain. In contrast, the blood vessels of the kidneys and the extremities did not become vasodilated, even at an arterial Po_2 as low as 20 torr. These changes in vascular tone would shunt more blood to the heart and the brain (49).

Local perfusion with angiotensin or epinephrine resulted in a similar redistribution of blood flow in response to hypoxia, except that it was due to selective vasoconstriction rather than vasodilation. For example, with an arterial oxygen tension of 40 torr, the normal vasoconstrictor response to angiotensin was maintained in skeletal muscle but it was abolished in the coronary vessels, favoring a redistribution of blood flow away from skeletal muscle and to the heart muscle during hypoxia (49).

Stimulation of *chemoreceptors* in the carotid bodies by hypoxia produces a neurally mediated group of vascular responses that result in a redistribution of blood flow that generally parallels the local vascular effects: there is vasoconstriction in skeletal muscle and intestine, either no response or vasodilation in the brain, and vasodilation in the coronary vessels. The end result, as with the local vascular effects, is a redistribution of blood to the heart and the brain, organs with high oxygen requirements (49).

The extent to which the acute changes in regional blood flow in adult animals are relevant to the infant with anemia of prematurity depends on (*a*) whether similar vascular responses occur at early stages of development; (*b*) whether they also occur under mild conditions of hypoxia; and (*c*) whether there is evidence of corresponding changes with more chronic hypoxia in premature infants. The first two points have been addressed in studies of fetal lambs near term in which radioactively labeled microspheres (50) were used to estimate organ blood flow at different levels of fetal oxygenation (51). As in a variety of experiments with adult animals, blood flow to the brain and heart muscle increased as arterial oxygen content decreased; as a consequence, the arterial oxygen supply remained nearly constant over a sixfold range in arterial oxygen content. This adaptation in blood flow to maintain a uniform oxygen supply was evident even with the mildest degrees of hypoxia. Blood flow to the kidney, digestive tract, and carcass tended either to remain relatively constant or to increase slightly with moderate hypoxia, but presumably not enough to compensate for the decrease in oxygen content; with severe hypoxia, there was a pronounced decrease in blood flow to all of these organs. These studies indicate that the near-term lamb fetus can redistribute blood flow in a manner similar to mature animals and that mild degrees of hypoxia can produce these effects. Furthermore, the restriction of oxygen supply to the kidney early in the course of hypoxia provides a mechanism for increased production of erythropoietin and increased erythropoiesis in response to hypoxia.

Although changes in distribution of blood flow are of major importance, they are difficult to evaluate in anemic infants. Most studies of the redistribution of blood flow with hypoxia have employed invasive techniques that cannot be applied to patients. Furthermore, even in the experimental animal, we know little about the harmful consequences of chronic hypoxia on the growth and function of tissues whose oxygen supply is diminished in order to preserve that of the heart and brain. Rather than attempt to evaluate altered regional blood flow directly in infants, it may prove more practical to detect it indirectly, by the measurement of erythropoietin, which is discussed next.

Erythropoietin in the Neonatal Regulation of Erythropoiesis

Red cell production is normally regulated by the glycoprotein hormone erythropoietin (Ep). The major site of Ep production, activation, or both is the kidney. However, the possibility of a different site of production during early development is suggested by data in the rat showing a shift from hepatic to renal production during the neonatal period (52). The amount of Ep produced depends not only on the concentration of hemoglobin but also on the affinity of hemoglobin for oxygen (34). Production of Ep is decreased in response to transfusion and is increased by anemia or hypoxia (53). The target of Ep is the erythroid precursor cell. Moderate concentrations of Ep stimulate primarily the proliferation of more mature and differentiated erythroid precursors, whereas high concentrations also stimulate the division of earlier, less differentiated cells (54). This stimulation eventually results in a rise in hemoglobin concentration to a level that eliminates the hypoxic stimulus and that allows Ep production to decrease to baseline rates.

If premature infants are subject to hypoxia as the hemoglobin concentration reaches its nadir, it implies an impairment in the regulation of erythropoiesis. There is no doubt that the fetus and newborn can produce Ep in a *qualitatively* appropriate fashion. Ep activity is detectable in cord blood (55), and elevated urinary Ep values are found in premature infants who are born anemic or in infants who are small-for-dates as a result of placental insufficiency or intrauterine hypoxia (56, 57). There is less information about postnatal changes in erythropoietin. Judging from changes in the percentage of reticulocytes in the peripheral blood (2) and erythroid precursors in the bone marrow (2, 23, 24), one would anticipate a marked decrease in erythropoietin shortly after birth, followed by a rise in values between 1 and 3 months after birth, when more active red cell production resumes. To some extent, these expectations are supported by experimental data. Plasma Ep measured by bioassay could not be detected after a few days of age (55, 58), but it became measurable again at 2 and 3 months of age (58). A more sensitive radiommunoassay yielded measurable values during early infancy in a group of premature infants. High Ep values correlated roughly with low hemoglobin concentration and with low oxygen-unloading capacity (38). Moreover, at comparable hemoglobin concentrations, infants with a low percentage of fetal hemoglobin and decreased oxygen affinity tended to have lower Ep values, usually as a result of either exchange or multiple replacement transfusion. All of these results indicate a rise in Ep under circumstances that would decrease the capacity to deliver oxygen to tissues. Thus, a rise in Ep appears to characterize the nadir of the anemia and is probably responsible for the subsequent rise in the concentration of hemoglobin.

It is reasonable to attribute a postnatal fall in hemoglobin concentration to improved oxygen delivery and a decreased production of Ep. However, it is more difficult to explain why hemoglobin falls to values lower in premature than in term infants. The possibility that the anemia of prematurity is *not* accompanied by hypoxia has been considered by several investigators (6, 9, 59). There is evidence that oxygen consumption (ml O_2/kg/min) and heat production (kcal/kg/24 hr), measured in a neutral thermal environment, are lower in the premature than in the term infant. The analogy is drawn to older patients with hypothyroidism in whom anemia is also believed to be due to decreased tissue requirements for oxygen. However, it is difficult to reconcile this line of reasoning with recent reports of a relatively high prevalence of clinical manifestations of hypoxia that are reversed by transfusion. An additional explanation for the exaggeration of the anemia is that the erythropoietic response to hypoxia is decreased in the premature infant compared to the older child or adult. This might be attributed either to a somewhat impaired production of erythropoietin or to a lowered responsiveness of erythroid precursor cells to a given erythropoietin concentration.

Previous studies of erythropoietin in premature infants have been hampered by lack of a practical and reliable assay. The bioassay is cumbersome, imprecise, and insensitive; values in the normal and subnormal range cannot be accurately measured in most procedures, and sera may contain nonspecific stimulators and inhibitors of erythropoiesis (34). Furthermore, the large amount of serum required for the assay has discouraged its use in studies of infants. Subsequently developed radioimmunoassays had the advantage of greater sensitivity but also had serious drawbacks. Ep was not sufficiently purified to assure specificity of the assay and in many cases the results were inconsistent with biologic activity. Recently, a radioimmunoassay based on a pure antigen has been developed (60). This assay may make it easier to determine whether anemia of prematurity is related to diminished production of Ep in response to a given hypoxic stimulus or to decreased responsiveness of red cell precursors to a given concentration of Ep.

Clinical Approaches to Transfusion Therapy

It can be appreciated from the preceding discussion that the regulation of erythropoiesis in the perinatal period is more complex than at any other time of life. The lack of specific clinical guidelines for the transfusion of premature infants is therefore not surprising.

Most transfusions for anemia of prematurity are administered in one of two situtations. The first, which to some degree represents prophylactic treatment, is the common practice of replacing blood after relatively large amounts are removed for laboratory studies. Such blood relacement is performed in infants who require frequent monitoring of blood gases or in whom more than an estimated 5–10% of the blood volume is removed over a brief period of time. Fewer blood transfusions are required if blood drawing is kept to the minimum required for good clinical management. This necessitates the use of microanalytic procedures and the elimination of laboratory studies that do not influence treatment. In any case, it is a good clinical routine to maintain an accurate record of the amount of blood removed. The special considerations that apply to transfusions in infants are the subject of a recent review (61). Among these is the use of special routines adapted to the administration of small amounts of blood and to the use of blood derived from as few donors as possible.

The second type of transfusion is commonly administered during the period when hemoglobin values are at their lowest, between 1 and 3 months after birth. In some infants this anemia of prematurity appears to be well compensated. Others are reported to develop apnea, tachycardia, poor feeding, and slow weight gain, which are reversed by blood transfusion (14–16). Ideally, one would like to anticipate such late manifestations of hypoxia but also to avoid transfusing infants whose tissues are adequately oxygenated despite low hemoglobin values.

A blood transfusion, particularly if given solely on the basis of the hemoglobin concentration, is likely to be the first of a series. The first transfusion will raise the P_{50} by increasing the concentration of hemoglobin A. This will facilitate oxygen delivery and then depress red cell production, which eventually decreases the concentration of hemoglobin. In addition to their mere temporary effectiveness, blood transfusions carry additional risks, among which are the transmission of hepatitis and cytomegalic inclusion disease (62, 63).

When considering the need for a transfusion, first compare the concentration of hemoglobin to age-specific reference values for preterm infants (Figure 1). Hemoglobin values normally decrease on the average of about 1 g/dl/week between 2 and 8 weeks of age (5, 6). Thus, a hemoglobin concentration of 9.5 at 2 weeks of age is markedly subnormal, whereas it is close to the mean for healthy premature infants 8 weeks after birth.

The reticulocyte count is also of value since it helps to anticipate a change in hemoglobin concentration. A slight elevation in the reticulocyte count in an otherwise healthy infant (Figure 1) suggests that the hemoglobin concentration is about to increase spontaneously. However, if the hemoglobin concentration does not rise despite an elevated reticulocyte count, the possibility of hemolytic disease must be considered; vitamin E deficiency, residual effects of hemolytic disease of the newborn, and hereditary spherocytosis are prominent in the differential diagnosis.

The infant's transfusion history is another important factor in interpreting the hemoglobin concentration. An infant who has had an exchange transfusion or has received a large volume of blood in the form of replacement transfusions will have an increased percentage of hemoglobin A and a rise in P_{50}. As a result of enhanced oxygen release to the tissues, such infants appear to equilibrate at a concentration of hemoglobin that averages 2–3 g/dl less than values in untransfused infants (38).

In addition to the concentration of hemoglobin, other laboratory tests that reflect on oxygen delivery are the P_{50} itself, or DPG and fetal hemoglobin which influence the P_{50}. Recently developed instruments allow the measurement of P_{50} on a few microliters of blood. The procedure is still relatively difficult, but improvements in technology should make the measurement of P_{50} more practical in intensive care nurseries. At any given concentration of hemoglobin, a high P_{50} would weigh more against transfusion than a low one.

Among the clinical manifestations, a consistent increase over earlier baseline values in the heart and respiratory rates suggests the possibility of hypoxia. A lower than expected rate of weight gain may also be a manifestation of hypoxia (16). Conceivably, the compensatory preservation of oxygen delivery to the brain and myocardium in response to hypoxia diminishes the supply of oxygen to other tissues; perhaps a decreased rate of growth is part of the price paid for this compensation.

Preventable Components of the Anemia

Some of the factors that can accentuate the anemia of prematurity are *iron deficiency, vitamin E deficiency,* and *frequent blood sampling* for laboratory studies. Premature infants often have cardiorespiratory problems requiring intensive care and repeated monitoring of blood gases and pH. Even with micro methods, such studies can remove more than 10% of the blood volume within a few days. Recently, methods have been developed for the transcutaneous monitoring of oxygen and carbon dioxide (64) and for the continuous measurement of oxygen saturation of hemoglobin by an arterial catheter oximeter (65). If these methods prove to be of as much clinical value as the removal of blood for the direct measurement of blood gases, they may drastically decrease the need for replacement transfusion.

Iron deficiency is not likely to play a role in the anemia of prematurity during the first 2 months of life (5) unless there has been perinatal blood loss or unusually large amounts of blood have been removed for laboratory studies. After 2 months, iron deficiency anemia is almost certain to develop in premature infants unless supplemental iron is administered. The anemia is most likely to manifest itself in a failure to increase the hemoglobin concentration substantially between 2 and 4 months of age. A dose of iron of about 2 mg/kg/day in the form of ferrous sulfate drops or iron-fortified formula should be adequate to prevent iron deficiency if started between 2 weeks and 2 months after birth (5, 66). Recently it was suggested that the dose of iron be increased to 4 mg/kg/day between 4 and 12 months of age, when iron needs are greatest (29).

Vitamin E deficiency, in contrast to iron deficiency, is likely to accentuate the anemia of prematurity during the first 2 months after birth. Vitamin E deficiency is more severe in infants with a birth weight under 1500 grams (67). The deficiency is due to a markedly depressed absorption of this fat-soluble vitamin and is spontaneously corrected after 2–3 months of age when the absorption of vitamin E improves (68). Lack of the vitamin results in a hemolytic anemia that is attributed to oxidative breakdown of the lipids in the red cell membrane. Iron functions as a cofactor in lipid peroxidation and aggravates hemolysis, particularly when given in large doses (67). The risk of hemolysis also depends on the degree of saturation of fatty acids in the diet. When the diet is relatively rich in polyunsaturated fatty acids, as with many infant formulas, the composition of the lipids in the red cell membrane undergoes a corresponding change, and the cell becomes more subject to oxidative damage and hemolysis (69). The interaction among these three factors—vitamin E deficiency, iron administration, and a diet rich in polyunsaturated fatty acids—became evident when small premature infants were fed certain iron-fortified proprietary infant formulas. Now that the problem has been recognized, vitamin E deficiency can be treated in one or more of three ways: (*a*) iron administration may be delayed (66); (*b*) a bioavailable, water-soluble form of vitamin E can be administered and can overcome the malabsorption of the naturally occurring vitamin (70); or (*c*) human milk or a formula comparably rich in saturated fatty acids can be fed to small premature infants (69).

Concluding Comments

The anemia of prematurity results from a rapid and profound decrease in the concentration of hemoglobin that occurs primarily in response to increased availability of oxygen after birth. This decline is compensated for, to varying degrees, by an increased capacity of hemoglobin to unload oxygen to the tissues. Why the concentration of hemoglobin falls to such extremely low values, particularly in the premature infant, is not understood, but could be related either to a lack of hypoxic stimulus due to low oxygen requirements or to an altered quantitative relationship between hypoxia, erythropoietin production, and the concentration of hemoglobin. If hypoxia develops, the essential functions of the heart and brain tend to be protected by a redistribution of blood flow, but this adaptation may occur at the expense of body growth and organ function.

Clinical management of the anemia of prematurity first requires a familiarity with the normal developmental changes that influence the delivery of oxygen. In addition, it is important to prevent those components of the anemia that are due to excessive blood drawing, iron deficiency, or vitamin E deficiency. Nevertheless, in some infants there will be evidence of compromised oxygen delivery and a blood transfusion will be indicated. Specific laboratory criteria for transfusion remain difficult to establish, but both hemoglobin concentration and oxygen affinity must be taken into account. In the future, the decision may be facilitated by the clinical application of new instruments that measure the oxygen dissociation curve of blood and by the emergence of a better assay for erythropoietin.

At present, the transfusion policy with respect to the anemia of prematurity varies widely and depends to a large degree on how the clinician views this condition. Viewing the anemia of prematurity as a normal developmental process discourages the use of transfusion as a form of tampering with a self-limited, physiologic phenomenon. Such conservatism is in the best tradition of clinical teaching and is partly a reaction to many examples of excessive therapeutic zeal: too much oxygen in premature infants leading to retrolental fibroplasia, irradiation of large thymus glands resulting in thyroid tumors, etc. This approach leans toward withholding transfusion unless there is a substantial weight of clinical and laboratory evidence in its favor; when transfusions are deemed necessary, there may also be an effort to administer relatively small amounts of blood so that the onset of spontaneous recovery is not delayed by prolonged suppression of marrow erythropoietic activity.

From an opposing point of view, one could argue that the anemia of prematurity is an exaggeration of a physiologic process and that it exists primarily by virtue of the recently improved survival of small premature infants, which was made possible only by intensive medical intervention. This argument is favored by recent studies indicating that clinical features of anemia were present in over half of infants who were <32 weeks gestational age at birth, and that improved weight gain commonly followed transfusion in small premature infants (16). If the results can be confirmed under carefully controlled conditions, it is likely that more and larger transfusions will be used in those premature infants of lowest birth weight.

At present, in weighing the risks of intervention against potential benefits, it seems safest to steer a middle course. One should be alert to the infant whose anemia is poorly compensated, but at the same time it is important to avoid the attribution of hypoxia to small clinical perturbations that are also characteristic of infants who have no problems with oxygen delivery. Fortunately, research in the area of oxygen delivery in the premature continues to be active, and firmer clinical guidelines will likely emerge before long.

Literature Cited

1. Gairdner, D., Marks, J., Roscoe, J. D. 1955. Blood formation in infancy. IV. The early anemia of prematurity. *Arch. Dis. Child.* 30:203–11
2. Seip, M., Halvorsen, S. 1956. Erythrocyte production and iron stores in premature infants during the first months of life. The anemia of prematurity—etiology, pathogenesis, iron requirement. *Acta Paediatr. Scand.* 45: 600–17
3. Schulman, I. 1959. Anemia of prematurity. *J. Pediatr.* 54:663–72
4. O'Brien, R. T., Pearson, H. A. 1967. Physiologic anemia of the newborn. *J. Pediatr.* 79:132–38
5. Lündstrom, U., Siimes, M. A., Dallman, P. R. 1977. At what age does iron supplementation become necessary in low birth weight infants? *J. Pediatr.* 91:878–83
6. Stockman, J. A. III 1977. Anemia of prematurity. *Clin. Perinatol.* 4:239–57
7. Delivoria-Papadopoulos, M., Roncevic, N., Oski, F. A. 1971. Postnatal changes in oxygen transport of term, premature, and sick infants: the role of red cell 2,3-diphosphoglycerate and adult hemoglobin. *Pediatr. Res.* 5:235–45
8. Versmold, H., Seifert, G., Riegel, K. P. 1973. Blood oxygen affinity in infancy: the interaction of fetal and adult hemoglobin, oxygen capacity and red cell hydrogen ion, and 2,3-diphosphoglycerate concentration. *Resp. Physiol.* 19:14–25
9. Delivoria-Papadopoulos, M. 1978. Neonatal oxygen transport. In *Reviews in Perinatal Medicine,* ed. E. M. Scarpelli, E. V. Cosim, 2:345–61. New York: Rowen. 395 pp.
10. Dawes, G. S., Mott, J. C. 1959. The increase in oxygen consumption of the lamb after birth. *J. Physiol. London* 146:295–315
11. Scopes, J. W., Ahmed, I. 1966. Minimal rates of oxygen consumption in sick and premature newborn infants. *Arch. Dis. Child.* 41:407–16
12. Hill, J. R., Robinson, D. C. 1968. Oxygen consumption in normally grown, small-for-dates, and large-for-dates new-born infants. *J. Physiol.* 199:685–703
13. Mott, J. C. 1961. The ability of young mammals to withstand total oxygen lack. *Br. Med. Bull.* 17:144–47
14. Wardrop, C. A. J., Holland, B. M., Veale, K. E. A., Jones, J. G., Gray, O. P. 1978. Nonphysiologic anemia of prematurity. *Arch. Dis. Child.* 53: 855–60
15. Stockman, J. A. III, Levin, E., Clark, D., Graeber, J. E. 1979. Oxygen consumption of premature infants in the first 10 weeks of life: the response to transfusion. *Pediatr. Res.* 13:442 (Abstr.)
16. Stockman, J. A. III, Clark, D. A., Levin, E. A. 1980. Weight gain—a response to transfusion in preterm infants. *Pediatr. Res.* 14:612 (Abstr.)
17. Oski, F. A., Naiman, J. L. 1966. *Hematologic Problems of the Newborn,* pp. 1–17. Philadelphia: Saunders. 400 pp.
18. Heymann, M. A., Rudolph, A. M. 1972. Effects of congenital heart disease on fetal and neonatal circulation. *Prog. Cardiovasc. Dis.* 15:115–43
19. Bratteby, L. E., Garby, L., Groth, T., Schneider, W., Wadman, B. 1968. Studies on erythrokinetics in infancy. XIII. The mean life-span frequency function of red blood cells formed during foetal life. *Acta Paediatr. Scand.* 57:311–19
20. Halvorsen, S., Finne, P. H. 1968. Erythropoietin production in the human fetus and newborn. *Ann. NY Acad. Sci.* 149:576–77
21. Finne, P. H. 1966. Erythropoietin levels in cord blood as an indicator of intrauterine hypoxia. *Acta Paediatr. Scand.* 55:478–88

22. Garby, L., Sjölin, S., Vuille, J. 1963. Studies on erythrokinetics in infancy. III. Disappearance from plasma and red cell uptake of radioactive iron injected intravenously. *Acta Paediatr. Scand.* 52:537–53
23. Shapiro, L. M., Bassen, F. A. 1941. Sternal marrow changes during the first week of life. Correlation with peripheral blood findings. *Am. J. Med. Sci.* 202: 341–54
24. Rosse, C., Kraemer, M. J., Dillon, T. L., McFarland, R., Smith, N. J. 1977. Bone marrow cell populations of normal infants: the predominance of lymphocytes. *J. Lab. Clin. Med.* 89: 1225–40
25. Pearson, H. A. 1967. Lifespan of the fetal red blood cell. *J. Pediatr.* 70: 166–71
26. Saarinen, V. M., Siimes, M. A. 1978. Developmental changes in red blood cell counts and indices of infants after exclusion of iron deficiency by laboratory criteria and continuous iron supplementation. *J. Pediatr.* 92:412–16
27. Dallman, P. R., Siimes, M. A., Stekel, A. 1980. Iron deficiency in infancy and childhood. *Am. J. Clin. Nutr.* 33:86–118
28. Saarinen, V. M. 1978. Need for iron supplementation in infants on prolonged breast feeding. *J. Pediatr.* 93: 177–80
29. Lundström, U., Siimes, M. A. 1980. Red blood cell values in low-birth-weight infants: ages at which values become equivalent to those of term infants. *J. Pediatr.* 96:1040–42
30. Bard, H., Teasdale, F., 1979. Red cell oxygen affinity, hemoglobin type, 2,3-diphosphoglycerate, and pH as a function of fetal development. *Pediatrics* 64: 483–87
31. Delivoria-Papadopoulos, M., Miller, L. D., Forster, R. E. II, Oski, F. A. 1976. The role of exchange transfusion in the management of low birth weight infants with and without severe respiratory distress syndrome. I. Initial observations. *J. Pediatr.* 89:273–78
32. Benesch, R., Benesch, R. E. 1967. The effect of organic phosphates from the human erythrocyte on the allosteric properties of hemoglobin. *Biochem. Biophys. Res. Commun.* 26:162–67
33. Chanutin, A., Curnish, R. R. 1967. Effects of organic and inorganic phosphates on the oxygen equilibrium of human erythrocytes. *Arch. Biochem. Biophys.* 121:96–102
34. Adamson, J. W., Finch, C. A. 1975. Hemoglobin function, oxygen affinity, and erythropoietin. *Ann. Rev. Physiol.* 37:351–69
35. Oski, F. A., Delivoria-Papadopoulos, M. 1970. The red cell, 2,3-diphosphoglycerate, and tissue oxygen release. *J. Pediatr.* 77:941–56
36. Ortalesi, M. M., Hay, W. W. 1971. Regulation of oxygen affinity of fetal blood. I. In vitro experiments and results in normal infants. *Pediatrics* 48:857–64
37. Bard, H. 1973. Postnatal fetal and adult hemoglobin synthesis in early preterm newborn infants. *J. Clin. Invest.* 52: 1789–95
38. Stockman, J. A. III, Garcia, J. F., Oski, F. A. 1977. The anemia of prematurity. Factors governing the erythropoietin response. *N. Engl. J. Med.* 296:647–50
39. Hilpert, P., Fleischmann, R. G., Kempe, D., Bartels, H. 1963. The Bohr effect related to blood and erythrocyte pH. *Am. J. Physiol.* 205:337–40
40. Duhm, J., Gerlack, E. 1974. Metabolism and function of 2,3-diphosphoglycerate in red blood cells. In *The Human Red Cell in Vitro,* ed. T. J. Greenawalt, G. A. Jamieson, pp. 111–48. New York: Grune & Stratton. 363 pp.
41. Lenfant, C., Torrance, J., English, E., Finch, C. A., Reynafarje, C., Ramos, J., Faura, J. 1968. Effect of altitude on oxygen binding by hemoglobin and organic phosphate levels. *J. Clin. Invest.* 47:2652–56
42. Jones, J. G., Holland, B. M., Veale, K. E. A., Wardrop, C. A. J. 1979. "Available oxygen," a realistic expression of the ability of the blood to supply oxygen to tissues. *Scand. J. Haematol.* 22: 77–82
43. Oski, F. A. 1973. Designation of anemia on a functional basis *J. Pediatr.* 83: 353–54
44. Lister, G., Walter, T. K., Versmold, H. T., Dallman, P. R., Rudolph, A. B. 1979. Oxygen delivery in lambs: cardiovascular and hematologic development. *Am. J. Physiol.* 237(6):H668–H675
45. Riggs, T. E., Shafer, A. W., Guenter, C. A. 1973. Acute changes in oxyhemoglobin affinity. Effects on oxygen transport and utilization. *J. Clin. Invest.* 52:2660–63
46. Bing, O. H. L., Keefe, J. F., Wolk, M. J., Lipana, J. G., McIntyre, K. M., Levine, H. J. 1969. Cardiovascular responses to hypoxia and varying PCO_2 in the awake dog. *J. Appl. Physiol.* 27:204–8

47. Finch, C. A., Lenfant, C. 1972. Oxygen transport in man. *N. Engl. J. Med.* 286:407–15
48. Lambertsen, C. J. 1965. Medical implications of high oxygen pressures. *Coll. Physicians Philadelphia Trans.* 33:1–18
49. Heistad, D. D., Abboud, F. M. 1980. Circulatory adjustments to hypoxia. *Circulation* 61:463–70
50. Rudolph, A. M., Heymann, M. A. 1967. The circulation of the fetus in utero. Methods for studying distribution of blood flow, cardiac output, and organ blood flow. *Circ. Res.* 21:163–84
51. Peeters, L. L. H., Sheldon, R. E., Jones, M. D. Jr., Makowski, E. L., Meschia, G. 1979. Blood flow to fetal organs as a function of arterial oxygen content. *Am. J. Obstet. Gynecol.* 135:637–46
52. Peschle, C., Marone, G., Genovese, A., Magli, M. C., Condorelli, M. 1976. Erythropoietin production in the rat: additive role of kidney and liver. *Am. J. Physiol.* 230:845–48
53. Schooley, J. C., Mahlmann, L. J. 1975. Hypoxia and the initiation of erythropoietin production. *Blood Cells* 1:429–48
54. Peschle, C. 1980. Erythropoiesis. *Ann. Rev. Med.* 31:303–14
55. Halvorsen, S. 1963. Plasma erythropoietin in cord blood and in blood during the first weeks of life. *Acta Paediatr. Scand.* 52:425–35
56. Matoth, Y., Zaizov, R. 1967. Absence of an inhibitor of erythropoiesis in postnatal plasma. *Isr. J. Med.* 3:477–79
57. Finne, P. H. 1968. Erythropoietin production in fetal hypoxia and in anemic uremic patients. *Ann. NY Acad. Sci.* 149:497–503
58. Mann, D. L., Sites, M. L., Donati, R. M., Gallagher, N. I. 1965. Erythropoietic stimulating activity during the first ninety days of life. *Proc. Soc. Exp. Biol. Med.* 118:212–14
59. Buchanan, G. R., Schwartz, A. D. 1974. Impaired erythropoietin response in anemic premature infants. *Blood* 44:347–52
60. Garcia, J. R., Sherwood, J., Goldwasser, E. 1979. Radioimmunoassay of erythropoietin. *Blood Cells* 5:405–19
61. Stockman, J. A. III, 1980. Red cell transfusion in the newborn: indications and unique blood banking needs. *Perspectives in Hematology and Oncology.* In press
62. Ballard, R. A., Drew, L., Hufnagel, K. G., Riedel, P. A. 1979. Acquired cytomegalovirus infection in preterm infants. *Am. J. Dis. Child.* 133:482–85
63. Spector, S. A., Schmidt, K., Ticknor, W., Grossman, M. 1979. Cytomegaloviruria in older infants in intensive care nurseries. *J. Pediatr.* 95:444–46
64. Whitehead, M. D., Pollitzer, M. J., Parker, D., Halsall, D., Delpy, D. T., Reynolds, E. O. R. 1980. Transcutaneous estimation of arterial PO_2 and PCO_2 in newborn infants with a single electrochemical sensor. *Lancet* 1:1111–14
65. Wilkinson, A. R., Phibbs, R. H., Gregory G. A. 1978. Continuous measurement of oxygen saturation in sick newborn infants. *J. Pediatr.* 93:1016–19
66. Committee on Nutrition, American Academy of Pediatrics 1976. Iron supplementation for infants. *Pediatrics* 58:765–68
67. Melhorn, D. K., Gross, S., Childers, G. 1971. Vitamin E–dependent anemia in the premature infant. I. Effects of large doses of medicinal iron. *J. Pediatr.* 79:569–80
68. Melhorn, D. K., Gross, S., Childers, G. 1971. Vitamin E–dependent anemia in the premature infant. II. Relationships between gestational age and absorption of vitamin E. *J. Pediatr.* 79:581–88
69. Williams, M. E., Shott, R. J., O'Neal, O. L., Oski, F. A. 1975. The role of dietary iron and fat on vitamin E deficiency anemia of infancy. *N. Engl. J. Med.* 292:887–90
70. Gross, S., Melhorn, D. K. 1974. Vitamin E–dependent anemia in the premature infant III. Comparative hemoglobin, vitamin E, and erythrocyte phospholipid responses following absorption of either water-soluble or fat-soluble d-alpha tocopheryl. *J. Pediatr.* 85:753–59

Ann. Rev. Med. 1981. 32:161–74

AMEBIC LIVER ABSCESS[1]

❖7389

R. S. Peters, M.D., N. Gitlin, M.D., and R. D. Libke, M.D.

Valley Medical Center, Veterans Administration Medical Center, Fresno, and
University of California School of Medicine, San Francisco, California 94143

INTRODUCTION

Entamoeba histolytica may cause diarrheal disease, acute or chronic abdominal or chest pain, cough, dyspnea fever, and weight loss. The potential for disease is present whenever one has been exposed to a lack of sewage facilities or lived in an endemic area. The clinical spectrum of amebiasis is broad, ranging from the asymptomatic passage of cysts through fulminant colitis to localized abscesses of the liver, lung, brain, or other tissues (74).

This brief review will focus on the amebic liver abscess. It includes current summaries of the epidemiology, etiology, pathogenesis, manifestations, diagnosis, and therapy.

EPIDEMIOLOGY

Amebiasis is found worldwide and *Entamoeba histolytica* is ubiquitous with its highest prevalence in the tropical and subtropical zones where sanitation is suboptimal. The prevalence in the United States 40 years ago was estimated at 10% of the population; the present incidence in the United States is less than 4% (11, 14, 18, 26). For more than 20 years an average of 3500 cases of amebiasis a year have been reported to the Center for Disease Control (CDC) in Atlanta, Georgia. These data represent information voluntarily submitted and probably underestimate the number of new cases each year (38). Investigations by the CDC indicate endemic foci of amebiasis exist in the United States; pockets of disease associated with poor sanitation have been reported from Indian reservations, from lower socioeconomic areas, and from mental institutions. Other groups at increased risk include immigrants and tourists from developing countries. The mode of transmission appears to be person-to-person rather than water contaminated with sewage containing amebic cysts (6). Venereal transmission of amebiasis has been reported and can result in genital, gut, or visceral disease (77).

---

[1] The US Government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this paper.

Liver abscess is the most common extraintestinal manifestation of amebiasis. Amebic abscess of the liver has been reported in 3–9% of all patients with amebiasis. A preponderance in males is noted, with the highest frequency in the 20–50-year age group (6, 12, 19, 36, 52, 62, 66). Reports from endemic areas indicate a significant incidence of amebic liver abscess in children under three years of age (31, 58). Visitors to tropical areas are more likely to develop invasive amebiasis than are permanent residents. Although most English and American series do not report racial differences (6, 19), in South Africa the blacks rather than the caucasians appear primarily affected (22, 75).

ETIOLOGY AND PATHOGENESIS

Two forms, trophozoites and cysts, of the protozoan parasite *Entamoeba histolytica* are found in stool specimens. The trophozoites are motile, rather small (10–20 μm) ameboid organisms with a single nucleus and a granular cytoplasm. This form is found in both symptomatic intestinal amebiasis with mucoid, bloody stools and in asymptomatic amebiasis. The nonmotile cyst form is similar in size to the trophozoite but may contain up to four nuclei. Infection occurs only with the ingestion of cysts because, unlike trophozoites, the cysts are resistant to gastric juice, to chlorine concentrations commonly used in domestic water purification, to room temperature, and to drying.

Finding trophozoites or cysts in the stool alone is not evidence of active disease. The trophozoite that causes invasive amebiasis may also live as a harmless commensal in the lumen of the bowel. Occasionally *E. histolytica* becomes invasive and causes varying degrees of intestinal, hepatic, and other disease. The factors that determine whether ingestion of *E. histolytica* cysts will produce either no infection, a commensal state, colitis, or other invasive disease are very poorly understood. Clinically this is reflected in patients with amebic liver abscess where only a minority of patients (16–33%) simultaneously have, or give a history of, active intestinal disease (1, 2, 4, 17, 19, 62).

Studies of Invasive Factors

A virulence-producing amebic-bacteria relationship has been shown in guinea pigs. Germ-free guinea pigs infected with *E. histolytica* did not develop invasive intestinal disease unless bacteria were present in the gut (54). Workers attempting to determine whether amebal viruses can convert the virulence of *E. histolytica* (46) have not found consistent changes. Amebal lytic enzymes were recently suggested as effectors of tissue damage. The measurement of the enzyme phosphoglucomutase has been offered as a likely laboratory method for distinguishing the pathogenic form, but this facet has not yet been completely resolved (60). The association of enzyme differences may not be intrinsic to the parasite but may reflect ingested bacteria.

The existence of two morphogically indistinguishable species might explain the presence of disease or the asymptomatic carrier state, as postulated by Brumpt half a century ago. He separated *E. dysenteriae,* a pathogen, from *E. dispar,* a more commonly occurring but harmless commensal (13). This view never gained acceptance but it has also never been disproved. The discovery of amebic cytotoxicity and of toxins in animal models has proved interesting but not altogether satisfying (9, 60). Thus, the nature of the trigger mechanism whereby ameba become invasive is still unclear. Host operative factors were discussed in a recent study in which susceptibility and resistance to *E. histolytica* could be demonstrated in embryonated eggs derived both from outbred heterogenous and from highly inbred chicks (32). Eggs from a genetically stable source were relatively constant in the distribution of susceptible and resistant embryos.

Hepatic Invasion

The latent period between the intestinal infection and the hepatic involvement is unexplained. The amebae are carried to the liver in the portal venous system, but the pathogenetic mechanisms for abscess formation are not fully known. The amebae apparently multiply and block small intrahepatic portal radicles, which causes focal infarction of hepatocytes; the organisms also produce a proteolytic enzyme that destroys the hepatocytes (65). Coalescence of the lesions into an abscess is assumed but not fully established. Solitary abscesses are the most common finding and occur especially in the right lobe superior-anteriorly, but multiple abscess formation of variable size and location in both lobes is not rare. The involvement of the left lobe and the presence of multiple abscesses are more common when the disease is advanced (19). Left lobe abscesses have the highest incidence of rupture especially into the pericardial or pleural sacs (5). Abscess cavity size varies from a few centimeters to 20 cm in diameter and can contain 1500 ml of fluid material. This material is classically described as anchovy paste, odorless and sterile; but it is often chocolate brown and bloody and may become secondarily infected, purulent, and gray-creamy in color (6). *E. histolytica* are seldom found in the fluid but their detection is enhanced if one examines the final drops aspirated rather than the initial aspirate (19). The amebae are limited to the zone of necrotic tissue surrounding the margin of the abscess and thus only drain from the abscess cavity once its volume and pressure decrease with aspiration. Amebic abscesses usually resolve without significant scarring.

Whether or not abscess formation is preceded by a phase of amebic hepatitis is debatable (23, 41, 43). Recent animal studies showed that six days after hepatic passage of *E. histolytica* the inoculated lobe had lesions containing multiple nidi of necrosis; six weeks later these lesions closely resembled advanced human amebic hepatic abscesses (43). No "hepatitis phase" could be identified. This is consistent with recent human pathological estimates (41, 62).

CLINICAL FEATURES

Symptoms

The usual onset and presentation of amebic liver abscess is with malaise and abdominal pain combined with hepatomegaly, fever, and weight loss over 4–30 days. This combination occurred in approximately 75% of cases in several series (6, 15, 63). Occasionally, low-grade chronic symptoms for months, even years, can be the presenting form (19). This variety is now relatively rare in the United States where the index of suspicion for amebiasis is more sensitive (6, 63). In our experience, amoebic liver disease may present initially with symptoms very suggestive of a respiratory infection. Patients may manifest cough, pleuritic chest pain, and dyspnea; only subsequently does the presence of abdominal symptoms become apparent. On occasion, patients may present acutely with symptoms simulating cholecystitis or appendicitis that suggest surgical intervention (37). Hepatomegaly in these cases may be the only significant factor suggesting amebiasis. Infrequently, presentation occurs at the time of rupture into the peritoneal cavity with features of an acute abdomen and shock (41). Rarely, signs of hepatic failure with encephalopathy and fever have been a form of initial presentation and, in these cases, hepatomegaly is also evident (59).

The symptoms of the colonic involvement or of an isolated liver abscess can become clinically active, more severe, or prolonged by the administration of corticosteroids for coexistent disease or by misdiagnosis of the colitic symptoms as due to nonspecific inflammatory colitis (70). The abdominal pain, usually the first symptom, is frequently described as a dull aching sensation, which on occasion has a stabbing characteristic, in the right upper quadrant, epigastric, or right lower chest. It was present in 88.6% of patients in a large series (19). The pain may increase with respiration or movement and is classically aggravated by gentle percussion over the liver (the "punch tenderness" sign). The pain may radiate to the right shoulder or laterally onto the chest wall, presumably depending on the location of the abscess in the liver. Left hepatic lobe abscesses may present with pain in the epigastrium or the left upper quadrant. Fever is present in 79–94% of cases (6, 19, 63). In most cases, the fever is persistent in the range of 38–39°C. High or spiking fevers are more often evident when the amebic abscess is secondarily infected (19, 62). Rigors or chills have been documented in approximately 30.4% of cases (12). Recent weight loss (as high as 60 pounds in the extreme) over 8–12 weeks has been observed in just over half the patients. The average weight loss is 5–20 pounds (6, 19, 62, 63). Malaise is prominent, and severe fatigue with almost total anorexia is found in many patients. There are conflicting reports regarding the frequency of cough and dyspnea. Cough was present in 30% and dyspnea in 11% of one series (15), but it was only mentioned as present in other reports (62, 63). Pleuritic chest pain occurs with a moderate degree of frequency, 12% of cases in one series (62). Diarrhea has been documented in 10–55% of patients at the time of presentation (6–15). Not all the diarrhea is of amoebic origin (62). Evidence of current amebic infestation of the bowel may be obtained by isolating the organism from the stools in 16–33% of cases (19, 63).

Physical Findings

Hepatomegaly is frequently encountered and is easily detectable. Enlargement to greater than 10 cm below the right costal margin was recorded in 10% of patients with advanced disease (15, 41). Percussion widths have not been recorded for large groups, but chest x rays of patients with amebic liver abscess have shown elevation of the right diaphragm in 70% of cases (41). The liver is generally smooth and diffusely tender. There is often a localized area of tenderness, presumably coinciding with an underlying superficial abscess, and "punch tenderness" to moderate percussion can be demonstrated. The hepatic consistency may be hard, simulating cirrhosis or hepatic carcinoma. Splenomegaly was detected in 14% of cases and ascites in 10% of cases in a series with advanced amebic liver disease (15). Hepatic bruits or friction rub are conspicuous by their absence, but the abscess may point through the abdominal wall in exceptional circumstances. Amoebic liver abscess in the cirrhotic liver is rare though both diseases may be very common in certain environments (25).

Jaundice occurs infrequently in amebic liver abscesses. Its presence is regarded as an ominous sign and is often associated either with an increased mortality (71) or with a large underlying amebic abscess, usually with a diameter in excess of 20 cm (62). Recently, a cholestatic type of jaundice was documented in patients with severe amebic liver abscess and, once again, this heralded a sinister outcome of the disease process (71).

Chest examination reveals rales, decreased breath sounds, dullness to percussion, and other features of an effusion or occasionally a friction rub usually on the right side. In one large series (41) elevation of the right diaphragm was present in 70%, pleural thickening in 27%, atelectasis in 14%, and a pleural effusion in 8% of cases. The findings occur on the left when that lobe is involved. Hemoptysis of fresh blood, pus, or anchovy paste-like material is seen only in patients with rupture of an abscess into the bronchi.

Progression

Secondary infection of an amebic abscess was documented in slightly more than 20% of patients in two series. Spiking or persistent high fever accompanied by leukocyte count in excess of 20,000/mm^3 should be regarded as strongly suggestive of this complication (19, 62). We believe bacterial infection must also be considered in all cases of sudden onset, rapid deterioration, or delayed response to therapy. The usual invading organisms are bowel flora. Pleural cavity extension occurs in 20% of cases (19). Penetration into the pleural cavity to produce empyema occurred in 7.9%, lung abscess or amebic pneumonitis in 5.6%, and bronchopleural fistula in 2% of cases (19, 41, 62). Intraabdominal extension following perforation into the peritoneal cavity is usually associated with shock and generalized peritonitis; it may be seen in as many as 7% of cases. The outcome of this complication varies (19, 41), ranging from uniformly fatal to recovery. This variance may be accounted for by the different patient populations. The outcome is poorer for malnourished than for well-nourished individuals. Rupture into the colon (41) or biliary tree (37) has been recorded in rare cases and so has the gradual development of a walled-off and localized subhepatic collection (19). The occurrence of separate coexisting amebic abscesses in the lung, brain, and spleen have been documented in cases of hepatic amebic abscesses, but these are extremely rare (19).

Left Hepatic Lobe Abscess

Left hepatic lobe abscess occurs far less frequently than the right lobe counterpart and accounts for only 4–18% of cases (5, 19, 41). Left lobe abscesses have a high incidence of rupture and, because of their anatomical site, frequently rupture or extend into the pericardial sac or left pleural cavity. A left lobe abscess should be strongly suspected if the pain is localized to the epigastrium or if a tender epigastric mass in present (5). Pericardial involvement by amebic disease appears to progress through three phases. Initially a "sympathetic effusion" appears when the pericardium and diaphragm become involved in an inflammatory reaction extending from the superior wall of a left lobe abscess. The second phase is rupture into the pericardial sac, and the third and later phase is development of constrictive pericarditis often following the successful treatment of frank perforation (5, 20, 41). Features that alert one to these complications are chest pain, pericardial friction rub, pericardial effusion, or signs of tamponade. Electrocardiographic changes are those of nonspecific pericarditis (20, 41). Drainage of the pericardial sac and the liver abscess, in addition to medical drug therapy, are imperative in order to secure a successful outcome. A well-timed subsequent pericardiectomy may be necessary (41). Early prompt aspiration of all left lobe amebic abscesses as well as medical therapy are advocated as a precaution against the pericardial complications.

DIAGNOSIS

Hematology and Biochemistry

Diagnosis will depend on showing structural changes of the liver parenchyma along with microscopic or serologic evidence of invasive amebic disease. Routine laboratory tests are usually nonspecific; the blood count may show a variable leukocytosis with white cell counts ranging from 5,000 to 33,000. One investigator found 18.3% of his patients with less than 9,000, 34% in a range 9,000–15,000, 37.4% ranging from 15,000–25,000, and 9.5% in excess of 25,000 (41). A leukocytosis of greater than 20,000/mm^3 is compatible with amebic liver abscess but should always arouse suspicion of either a primary pyogenic abscess or secondary bacterial infection. Mild to moderate anemia occurs in about 50% of patients (41, 62, 63). Hemoglobin of less than 8 g/dl is unusual (62) and is more likely in cases with longer duration of clinical symptoms. Eosinophilia is not a feature of amebic liver abscess. A markedly raised erythrocyte sedimentation rate with a mean level of 93 mm/hr was reported in a large series (17). Transaminase elevation is seen in less than half of the patients presenting with hepatic amebic abscesses. The SGOT level ranges between two and six times normal, with an average of two times. The SGPT elevation runs parallel to the SGOT. Alkaline phosphate was elevated in 69–80% of patients (6, 63) with a range of one- to fourfold increases over normal. The serum albumin is decreased in most patients, with levels as low as 1.9 g/dl (63). Moderate elevation of globulin is also frequently encountered. Bilirubin elevation is modest and is usually in the range of 2.5 mg/dl or less (63). There have been reports of bilirubin elevated as high as 4–10 mg/dl (62). A high bilirubin makes one suspect a complicated amebic abscess, impending hepatic failure (59, 71), or an extremely large abscess greater than 20 cm (62).

Chest Roentgenogram

The chest x ray frequently shows elevation of the right diaphragm (as already discussed). Pleural effusion with or without right parenchymatous lung densities is also frequently noted (41). Left lobe abscesses may cause a pericardial effusion or left pleural cavity reaction (5).

Demonstration of Structural Liver Defects

Liver scanning with radioactive colloids has been used to identify space-occupying liver lesions. Various types of isotopes have been used including ^{198}Au (64), ^{99m}Tc (35), ^{131}I (29), and ^{113}In (8); all seem to give satisfactory results. The most commonly used agent currently is ^{99m}Tc sulfurcolloid, probably because it gives adequate imaging with the least radioactivity. Lesions 2 cm or greater in size can usually be detected. Gallium scans with ^{67}Ga have demonstrated enhanced uptake of the liver tissue surrounding amebic liver abscesses, but the abscess cavities have not shown uptake of the radioactive gallium (27). The use of gallium in the detection of amebic liver abscesses has not been well studied and the sensitivity of this scanning technique has not yet been determined.

Demonstration of defects by isotope liver scanning does not distinguish between cystic or solid lesions; the use of ultrasound scanning gives additional support to the diagnosis of abscess and it can distinguish defects due to tumors. Hepatic amebic abscesses are less echogenic than normal liver and distal sonic enhancement is often seen. Over half of the abscesses show a homogenous pattern with low level echos. In an ultrasound study by Ralls et al (56), 41 of 42 abscesses were contiguous with the capsule. Ultrasound localization is also useful as a guide for aspiration of abscess contents. Computerized tomography can successfully demonstrate amebic liver abscess, but its value over other methods is subject to review (42).

Demonstration of Organism

MICROSCOPIC As already mentioned, stool examination for *E. histolytica* cannot be relied upon for diagnosis of an amebic liver abscess. In addition, the finding of cysts or trophozoites in the stool lends support to the clinical diagnosis of a hepatic defect as amebic, but does not confirm the diagnosis. Aspirates of amebic abscess contents contain trophozoites in 90% of cases when the last few milliliters aspirated are examined in a fresh state by experienced personnel, but this high rate of positivity is not universally reported (21). In addition to the examination for trophozoites, aspirated material should be gram-stained and cultured for aerobic and anaerobic bacteria routinely. Though demonstration of trophozoites in abscess contents is the best proof of diagnosis, aspiration for diagnosis alone is seldom required since serological tests give sufficient evidence of etiology in the majority of patients having amebic liver abscess.

SEROLOGY Indirect hemagglutination (IHA) is a complex test to perform but it is perhaps the most sensitive of the serologic tests available: 90–95% of patients with amebic liver abscesses have titers of at least 1:128 and a majority have titers of 1:16,000 or greater. This test also gives positive results in patients with amebic colitis without liver defects (especially in patients with chronic symptoms) and in a small percentage of patients with no demonstrable amebic disease (39, 47, 53, 69).

Gel diffusion precipitin tests are suitable for single specimen testing (IHA is not). This test is positive in 85–95% of patients with amebic liver abscess, but is also positive in 85–95% of patients with other diseases and results in up to 10% false positivity in normal blood donors residing in an endemic area (33, 40, 44, 53, 55). Patients who present soon after onset of symptoms may have negative findings on gel diffusion initially, but often convert to positive within a few days.

A latex agglutination test is also available and is simpler and faster than the gel diffusion test. It can be performed by relatively inexperienced personnel and results can be available within 10 minutes. It was found to give positive results in 98% of patients with amebic liver abscess, but was also positive in 15% of general medical patients and 5% of blood donors without evidence of amebic disease (48, 49, 50).

Counterimmunoelectrophoresis (CIE) is positive in more than 90% of sera from patients with extraintestinal amebiasis. It can be performed more rapidly than the gel diffusion tests and, with the differential reading of multiple band patterns, may distinguish patients with active disease from those with previous amebic infection (3, 61).

The cellulose acetate diffusion test is also sensitive and may be performed on test material that has been stored for extended periods (67). Complement fixation is less sensitive than most other methods of serologic testing. It requires considerable expertise and is not suitable for single specimen testing (33, 40, 68).

THERAPY

Medical Therapy

Medical therapy for amebic liver abscess requires either a single agent, or, alternatively, a combination of drugs that will eliminate the extraluminal parasites as well as the bowel organisms.

Nitroimidazoles, including metronidazole, are effective amebicides against intraluminal and extraluminal organisms. Metronidazole is effective against amebic liver abscess in over 90% of patients when doses of 750 mg three times per day are used. The duration of therapy is somewhat unclear but should continue for at least ten days (16, 24). Because of reports of relapse with the ten-day therapy, some authorities administer this agent for up to three weeks (24, 28, 73). Mild gastrointestinal symptoms, paresthesias, vertigo, ataxia, and leukopenia, have been described with metronidazole therapy but are reversible. Carcinogenic effects in rodents, mutagenic defects in bacteria, and teratogenicity in animals have been reported; however, no human carcinogenesis has been reported. The drug is available only for oral administration, but an intravenous preparation is currently being studied.

Tinidazole has not been used widely but has actions similar to metronidazole. The dose is 1.2 g per day for seven days, but for amebic liver abscess this dosage has not been firmly established (72). Metronidazole is generally available throughout the world whereas tinidazole is not widely available and has not been approved for use in the United States.

Chloroquine phosphate is effective against ameba residing in tissues and reaches high concentrations in liver tissue. It is administered gradually in a dose of 500 mg twice daily for two days followed by 500 mg each day for an additional 2–8 weeks. Patients treated for the more extended periods have a response rate of greater than 90% (16). Chloroquine is ineffective against intestinal disease. It may cause mild headache, itching, incoordination, peripheral neuritis, seizures, diarrhea, and ECG changes in a small number of patients. Retinopathy with long-term administration has been reported and deafness has been described in children of mothers who have taken this medicine during pregnancy.

Emetine hydrochloride is also an effective agent for extraintestinal amebiasis. Use of this agent requires hospitalization and monitoring for cardiotoxicity. In large or cumulative doses cardiac failure may result from focal necrosis of cardiac muscle. Dihydroemetine is thought to be less cardiotoxic but monitoring must still be performed. These agents are given in a dose of 1 mg/kg/day intramuscularly or by deep subcutaneous injection for 6–8 days.

Since chloroquine and emetine or dihydroemetine are ineffective against intestinal parasites, therapy with these agents must be combined with an intraluminal amoebicide. Diiodohydroxyquin (diodoquin) is most effective in the asymptomatic-cyst-carrier patients. Optic atrophy and blindness have been reported after prolonged use; rashes, gastrointestinal upset, and thyroid enlargement have been described as well (51). The usual dose is 650 mg three times a day orally for 20 days. Diloxanide furoate (furamide) is active against intestinal *E. histolytica* only in the presence of bacteria. The compound is safe and well tolerated with occasional gastrointestinal side effects (76). This drug is currently available only through the CDC in the United States.

Aspiration or Drainage

In some areas of the world, needle aspiration of hepatic amebic abscesses is routinely performed along with medical therapy (29); the need for this is questionable in view of the available effective pharmacological agents (57, 64). Medical therapy alone is adequate, especially in patients who are not toxic or distressed. Patients who do not respond to medical therapy within 48–72 hours should be considered candidates for needle aspiration or surgical drainage of their abscess (21). In addition, patients with impending rupture as evidenced by increasing pleural reaction, progressive diaphragmatic elevation, and increasing liver tenderness and enlargement require evacuation of their abscesses. Most investigators feel that patients with left lobe abscesses should have early drainage procedures performed in order to avoid extension into the pericardial space (33, 45, 72).

Total resolution time of amebic abscesses, with or without evacuation of abscess contents, varies from one to twelve months with the majority showing resolution by four months (30, 34). The total resolution time, as determined by radioisotope scanning, correlates poorly with the initial size of the abscess cavity. The persistence of an hepatic defect on scanning after effective therapy may not be associated with clinical or chemical evidence of liver disease. Therapeutic trials are sometimes useful in substantiating a clinical diagnosis of amebic liver abscess. When a therapeutic trial is used, chloroquine with or without emetine is preferred to metronidazole since patients with anaerobic liver abscesses may also show initial improvement with metronidazole therapy (37).

Literature Cited

1. Adams, E. B. 1976. The management of invasive amoebiasis. *S. Afr. J. Hosp. Med.* 53:327–35
2. Adams, E. B., MacLeod, I. N. 1977. Invasive amebiasis. II. Amebic liver abscess and its complications *Medicine* 56:325–34
3. Aguilar-Torres, F. G., Ross, E. J., Niebojewski, R. A., Jackson, L. J., Healy, G. R., Rytel, M. W. 1977. Comparison of counterimmunoelectrophoresis (CIE) with other serologic tests in the diagnosis of invasive amebiasis. *Abstr. Intern. Cong. Antimicrob. Agents Chemother.* II435
4. Aikat, B. K., Bhusnurmath, S. R., Chhuttani, P. N., Datta, D. V. 1978. Amoebic liver abscess—a clinicopathological study. *Indian J. Med. Res.* 67:381–91
5. Alkan, W. J., Kalmi, B., Kauderon, M. 1961. The clinical syndrome of amebic abscess of the left lobe of the liver. *Ann. Intern. Med.* 55:800–13
6. Barbour, G. L., Juniper, K. Jr. 1972. A clinical comparison of amebic and pyogenic abscess of the liver in sixty-six patients. *Am. J. Med.* 53:323–31
7. Berkman, J. M., Bargen, J. A. 1942. Amebic abscess of the liver with

choledochal and external fistulas: report of case. *Proc. Mayo Clin.* 17:483–87

8. Bhasin, A. S., Mittal, K. P., Shrivastava, P. N., Bhardwaj, O. P., Sachdew, S., Gupta, P. S. 1977. Resolution time of amoebic liver abscess by micro-opaque barium sulfate. *J. Trop. Med. Hyg.* 80:169–72
9. Bos, H. J. 1979. Toxins of *Entamoeba histolytica. Trans. R. Soc. Trop. Med. Hyg.* 73:120
10. Boultbee, J. E., Simjee, A. E., Rooknoodien, F., Engelbredt, H. E. 1979. Experiences with grey scale ultrasonography in hepatic amoebiasis. *Clin. Radiol.* 30:683–89
11. Brooke, M. M. 1964. Epidemiology of amebiasis in the U.S. *J. Am. Med. Assoc.* 188:519–21
12. Brown, P. W., Hodgson, C. H. 1938. Late results in treatment of amebic abscess and hepatitis of the liver. *Am. J. Med. Sci.* 196:305–447
13. Brumpt, M. E. 1925. Etude Sommare de l' *Entomeba dispar* N. Sp. Amibe a kystes quadrinueles, Parasite de l' Homme. *Bull. Acad. Med.* Paris. 94:943–52
14. Burrows, R. B. 1961. The prevalence of amebiasis in the United States and Canada. *Am. J. Trop. Med. Hyg.* 10:172–84
15. Chaves, D. C. J. Z., Cruz, I., Gomes, C., Domingues, W., DaSilva, E. M., Veleso, F. T. 1977. Hepatic amebiasis, analysis of 56 cases. *Am. J. Gastroenterol.* 68:134–40
16. Cohen, H. G., Reynolds T. B. 1975. Comparison of metronidazole and chloroquine for the treatment of amoebic liver abscess. *Gastroenterology* 69: 35–41
17. Cowie, R. L., Hickman, R., Saunders, S. J. Terblanche, J. 1972. Amoebic liver disease in Cape Town. *S. Afr. Med. J.* 46:1917–20
18. Craig, C. F. 1944. *The Etiology, Diagnosis and Treatment of Amebiasis,* pp. 45–50. Baltimore: Williams & Wilkins
19. DeBakey, M. E., Ochsner, A. 1951. Hepatic amebiasis: A 20 year experience and analysis of 263 cases. *Int. Abstr. Surg.* 92:209–31
20. Edwards, M. L. 1947. Amebic pericarditis. *Med. J. Aust.,* pp. 177–78
21. Eggleston, F. C., Verghese, M., Handa, M. B., Gill, S. S. 1978. The results of surgery in amebic liver abscess: experiences in eighty-three patients. *Surgery* 83:536–39
22. Elsdon-Dew, R. 1964. Amoebiasis. *Exp. Parasitol.* 15:87–96
23. Elsdon-Dew, R. 1969. Amoebiasis: Its meaning and diagnosis. *S. Afr. Med. J.* 43:483–87
24. Everett, E. D. 1974. Metronidazole and amebiasis. *Dig. Dis.* 19:626–36
25. Falaiye, J. M., Okeke, G. C. E., Fregene, A. O. 1980. Amebic abscess in the cirrhotic liver. *Gut* 21:161–63
26. Faust, E. C. 1942. The prevalence of amebiasis in the western hemisphere. *Am. J. Trop. Med.* 22:93–105
27. Geslien, G. E., Thrall, J. H., Johnson, M. C. 1974. Gallium scanning in acute hepatic amebic abscess. *J. Nucl. Med.* 15:561–63
28. Griffin, F. M. Jr . 1973. Failure of metronidazole to cure hepatic amebic abscess. *N. Engl. J. Med.* 288:1397
29. Habibullah, C. M., Quraishi, M. A., Chandra, V., Padmanabhan, Kumar, K., Agarwal, D. 1977. Liver scan in amoebic liver abscess. *J. Assoc. Physicians India* 25:593–96
30. Habibullah, C. M., Quraishi, M. A., Rao, S. R., Chandra, V., Murthy, K. J. R. 1977. Resolution of amebic liver abscess by serial liver scanning. *J. Indian Med. Assoc.* 69:250–52
31. Harrison, R. H., Crane, C. P., Fulginitis, V. A. 1979. Amebic liver abscess in children: clinical and epidemiologic features. *Pediatrics* 64:923–28
32. Jaouni, K. C. 1979. *Entamoeba histolytica:* genetic control of susceptibility in chicken eggs. *Exp. Parasitol.* 47: 54–64
33. Juniper, K., Worrell, C. L., Minshew, C., Roth, L. S., Cypert, H., Lloyd, R. E. 1972. Serologic diagnosis of amebiasis. *Am. J. Trop. Med. Hyg.* 21:157–68.
34. Kane, J. G., Fossieck, B. E. Jr., Parker, R. H. 1976. Metronidazole and hepatic abscess—a false positive response. *J. Am. Med. Assoc.* 236:2653–54
35. Kew, M. C., Osler, H. I., McCann, W. G., Esser, J., Levin, J. 1979. Radiocolloid imaging in primary liver cancer and amoebic liver abscess. *S. Afr. Med. J.* 56:127–29
36. Klatskin, G. 1946. Amebiasis of the liver: classification, diagnosis and treatment. *Ann. Intern. Med.* 25:601
37. Krettek, J. E., Goldstein, L. I., Busattil, R. N. 1979. The symptoms of an amebic abscess of the liver simulating an acute abdomen. *Surg. Gynecol. Obstet.* 148: 552–56
38. Krogstad, D. J., Spencer, H. C. Jr., Healy, G. R., Gleason, N. N., Sexton, D. J., Herron, C. A. 1978. Amebiasis: epidemiologic studies in the United

States, 1971–1974. *Ann. Intern. Med.* 88:89–97

39. Krupp, I. M. 1970. Antibody responses in intestinal and extraintestinal amebiasis. *Am. J. Trop. Med.* 19:57–62
40. Krupp, I. M., Powell, S. J. 1971. Antibody response to invasive amebiasis in Durban, South Africa. *Am. J. Trop. Med. Hyg.* 20:414–20
41. Lamont, N. McE., Pooler, N. P. 1958. Hepatic amoebiasis: A study of 250 cases. *Q. J. Med. New Ser.* 27:389–412
42. Levy, J. M., Nykamp, P. W., Jogerst, G., Yerger, F. S., Pitha, N. R., McFarland, J. O. 1978. CT-guided percutaneous drainage of an amoebic liver abscess. *Am. J. Gastroenterol.* 70:298–301
43. Lushbaugh, W. B., Kairalla, A. B., Pittman, F. E. 1977. Animal model for amebic liver abscess. *Gastroenterology* 72:1091
44. Maddison, S. E., Powell, S. J., Elsdon-Dew, R. 1965. Comparison of hemagglutinins and precipitins in amebiasis. *Am. J. Trop. Med. Hyg.* 14:551–53
45. Mandal, A. K., Thodepalli, H. 1978. Surgical aspects of amoebiasis and diagnostic clues. *Am. Surg.* 44:564–70
46. Mattern, C. F. T., Keister, D. B., Diamond, L. S. 1979. Experimental amebiasis. IV. Amebal viruses and the virulence of *Entamoeba histolytica. Am. J. Trop. Med. Hyg.* 28:653–57
47. Mehta, A. P., Deodhar, L. P., Gujarathi, P. S. 1978. Indirect hemagglutination test in intestinal and extraintestinal amoebiasis. *J. Postgrad. Med.* 24:103–5
48. Monroe, L. S., Korn, E. G., Fitzwilliams, S. J. 1972. Comparative study of latex agglutination and gel diffusion precipitin test in diagnosis of amebic liver abscess. *Am. J. Gastroenterol.* 58: 52–57
49. Morris, M. N., Powell, S. J., Elsdon-Dew, R. 1970. A rapid latex agglutination test for invasive amoebiasis. *S. Afr. Med. J.* 44:594–95
50. Morris, M. N., Powell, S. J., Elsdon-Dew, R. 1970. Latex agglutination test for invasive amoebiasis. *Lancet* June: 1362–63
51. Oakley, G. P. 1973. The neurotoxicity of the halogenated hydroxyguindines. *J. Am. Med. Assoc.* 225:395–97
52. Ochsner, A., DeBakey, M. E., Kleinsasser, R., DeBakey, E. 1942. Amebic hepatitis and hepatic abscess. *Rev. Gastroenterol.* 9:438–87
53. Patterson, M., Healy, G. R., Shabot, J. M. 1980. Serologic testing for amoebiasis. *Gastroenterology* 78:136–41
54. Phillips, B. P., Wolfe, P. A., Rees, C. W., Gordon, H. A., Wright, W. H., Reyniers, J. A. 1955. Studies on the ameba-bacteria relationship in amebiasis. *Am. J. Trop. Med.* 4:675–92
55. Powell, S. J., Maddison, S. E., Wilmet, A. J., Elsdon-Dew, R., 1965. Amoebic gel-diffusion precipitin test. *Lancet* Sept:602–3
56. Ralls, P. W., Meyers, H. I., Lapin, S. A., Rogers, W., Boswell, W. D., Halls, J. 1979. Gray-scale ultrasonography of hepatic amoebic abscesses. *Radiology* 132:125–29
57. Reynolds, T. B. 1971. Amebic abscess of the liver. *Gastroenterology* 60:952–54
58. Rode, H., Davies, M. R. Q., Cywes, S. 1978. Amoebic liver abscess in infancy and childhood. *S. Afr. J. Surg.* 16: 131–38
59. Saltzman, D. A., Smithline, N., Davis, J. R. 1978. Fulminant hepatic failure secondary to amebic abscesses. *Dig Dis.* 23:561–67
60. Sargeaunt, P. G., Williams, J. E., Grene, J. D. 1978. The differentiation of invasive and non-invasive *Entamoeba histolytica* by isoenzyme electrophoresis. 1978. *Trans. R. Soc. Trop. Med. Hyg.* 72:519–21
61. Savanat, T., Wanpen, C. 1969. Immunoelectrophoresis test for amoebiasis. *Bull. WHO* 40:343–53
62. Sepulveda, B., Jinich, H., Bassols, F., Monoz, R. 1959. Amebiasis of the liver: Diagnosis, prognosis and treatment. *Am. J. Dig. Dis.* 4:43–64
63. Shabot, M. J., Patterson, M. 1978. Amebic liver abscess: 1966–1976. *Dig. Dis.* 23:110–18
64. Sheehy, T. W., Parmleg, L. F., Johnston, G. S., Boyce, H. W. 1968. Resolution time of an amoebic abscess. *Gastroenterology* 55:26–34
65. Sherlock, S. 1968. *Diseases of the Liver and Biliary System.* Oxford, Edinburgh: Blackwell Scientific. 595 pp. 4th ed.
66. Smith, C., Ruffin, J. M. 1946. Amoebic infection of the liver as seen in North Carolina. *Gastroenterology* 6:294–97
67. Sodeman, W. A. Jr., Dowda, M. C. 1973. Rapid serologic methods for the demonstration of *Entamoeba histolytica* activity. *Gastroenterology* 65:35–41
68. Stamm, W. P., Phillips, E. A. 1977. A cellulose acetate membrane precipitin (CAP) test for amoebiasis. *Trans. R. Soc. Trop. Med. Hyg.* 71:490–92
69. Stein, J., Stein, B., Steinits, J., Talis, B. 1971. The indirect hemagglutination

test as an aid in the diagnosis of amebiasis. *Isr. J. Med. Sci.* 7:450

70. Stuiyer, P. C., Goud, J. L. M. 1978. Corticosteroids and liver amebiasis. *Br. Med. J.* 2:344–95
71. Tim, L. O. U., Segal, I., Hodkinson, H. I. 1978. Amoebic liver abscess in patients presenting with jaundice. *S. Afr. Med. J.* 55:179–84
72. Tschl-hyvn, J. 1974. Use of tinidazole (Fasigyn) on treatment of intestinal and hepatic amebiasis. *Proc. 3rd Int. Cong. Parasitol.* 3:1279
73. Wilde, H. 1973. Hepatic amebic abscess not responding to metronidazole. *N. Engl. J. Med.* 289:378
74. Wilmot, A. J. 1962. *Clinical Amoebiasis,* pp. 1–14. Oxford: Blackwell Scientific
75. Wilmot, A. J. 1962. See Ref. 74, p. 162
76. Wolfe, M. S. 1973. Nondysenteric intestinal amebiasis—treatment with diloxanide furoate. *J. Am. Med. Assoc.* 224: 1601–4
77. Vlvisaker, J. T., McDonald, G. B. 1980. Sexually acquired amebic colitis and liver abscess. *West. J. Gastrointest. Med.* 132:153–57

Ann. Rev. Med. 1981. 32:175–94
Copyright © 1981 by Annual Reviews Inc. All rights reserved

RELATIONS BETWEEN CIRCULATORY AND METABOLIC CHANGES IN SEPSIS

❖7390

John H. Siegel, M.D.

Departments of Surgery and Biophysical Sciences, State University of New York at Buffalo; Department of Surgery, The Buffalo General Hospital, 100 High Street, Buffalo, New York 14203

Introduction

The classic approach to human sepsis focuses on the role of various pathologic bacteria and fungi as infectious agents of varying potency. This review emphasizes that sepsis is an acquired disease of intermediary metabolism induced by infectious agents and producing characteristic patterns of metabolic and physiologic host responses.

These fundamental abnormalities in the biochemical processes and hormonal regulations of substrate utilization and oxidative energy generation interfere with the normal flux of energetic fuels and metabolic catabolic byproducts between muscle, liver, and adipose tissue. These fuel-energy deficits result in incomplete catabolism of aromatic amino acids and may generate as yet incompletely identified vasoactive substances that induce the pathophysiologic peripheral vascular pressure:flow relations characteristic of the septic process. In the septic host who retains the capability for a heightened sympathetic response, the resultant cardiac inotropic compensation to these vascular tone abnormalities appears to be a major factor in the development of the septic hyperdynamic state and its consequent pulmonary perfusion maldistribution abnormality. All of these factors interact with the septic-induced alterations in capillary permeability and with the energetically induced protein-calorie malnutrition to produce a septic autocannibalism which alters the balance between protein synthesis and protein catabolism. This enhanced protein catabolism and consequent urea synthesis may also be responsible for the decreased synthesis of reparative and immunologic proteins, which characterizes the later stages of the septic

0066-4219/81/0401-0175$01.00

response to injury, when a deficit in oxidative metabolism occurs. The septic patient then fails to utilize administered nutritional support, experiences a sequence of multiple organ failures producing myocardial, hepatic, pulmonary, and renal failure, and develops anergy and inadequate wound healing.

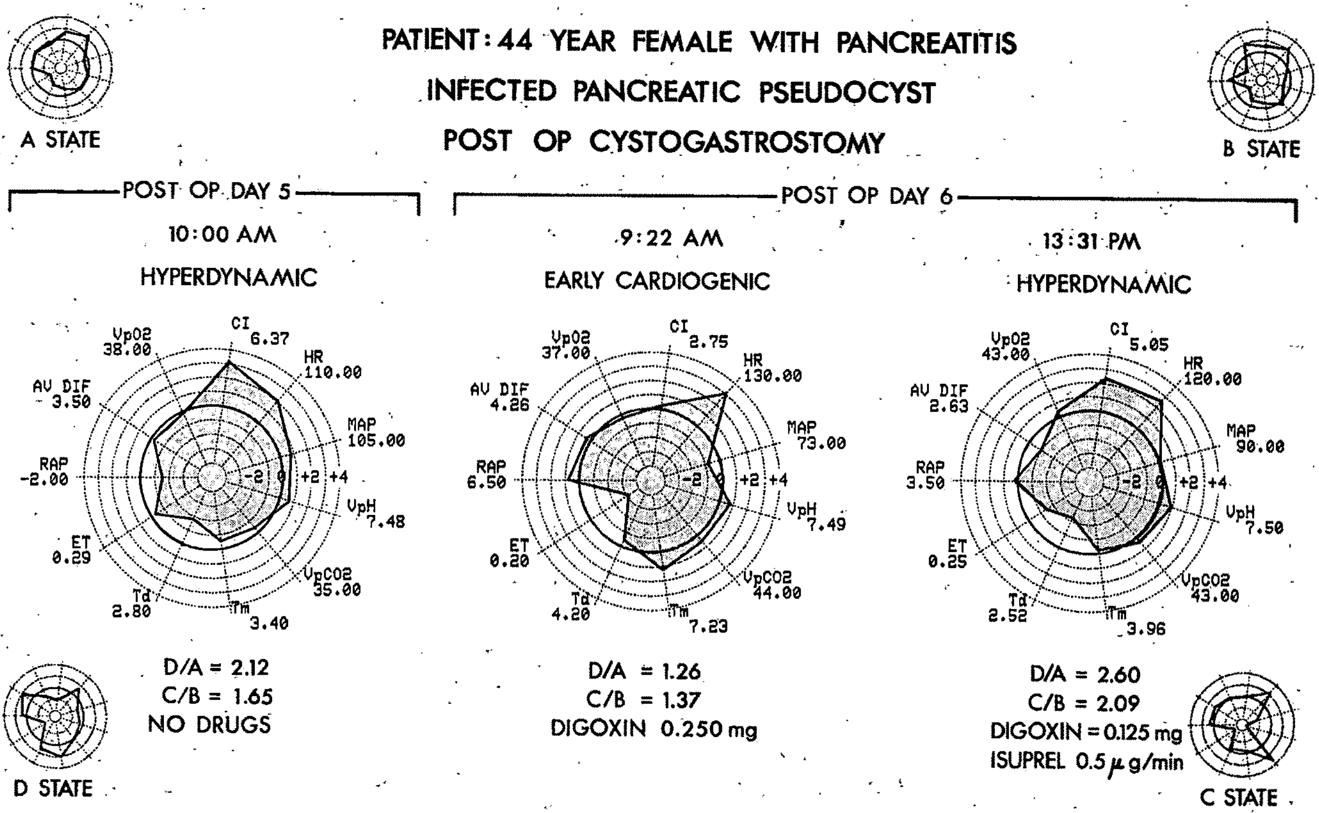
Patterns of Response and Classification

The physiologic and metabolic interrelations in sepsis appear independent of the type of infecting bacterial or fungal agent (1). However, infectious organisms of greater pathogenicity and enhanced antibiotic resistance, or infections appearing in an already compromised host, are more likely to induce the systemic septic response and to carry it on to the stage of metabolic decompensation. It is useful to consider the pattern of the physiologic adaptation to sepsis as an organization of host defenses from whose adequacy we may infer, and thus systematically classify, the severity of the underlying metabolic abnormalities (2). While this method of classification is obviously not the only possible approach to the problem (3–5), it is clinically useful since it can be carried out at the bedside with relative ease and good quantitative precision. It permits a strong clinical correlation with both the nature and magnitude of the observed biochemical abnormalities and with the probability of a fatal outcome if not corrected by medical or surgical measures. It also allows for the quantification of patient response to a therapeutic maneuver in real time and thus permits the effective direction of therapy.

Quantifying the pattern of physiologic abnormalities seen in the septic patient was attempted by Siegel et al (6, 7) and Friedman et al (8) who applied principles of statistical pattern recognition to a simultaneous set of multivariable physiologic data obtained from patients with various forms of shock and critical illness, including patients with sepsis and those with nonseptic traumatic injury or cardiogenic decompensation. They developed a frame of reference and a metric or scale by which abnormal pathophysiologic patterns could be discriminated from normal adaptive stress responses to injury. The patients' pattern of response could thus be quantified from one time period to the next. The data set used (Figure 1) included cardiac index (CI), heart rate (HR), mean arterial pressure (MAP), mean right atrial pressure (RAP), systolic ejection time (ET), the arteriovenous oxygen difference (AVo_2Dif), and the mixed venous Po_2 (Vpo_2), Pco_2 ($Vp\,co_2$), and pH (VpH). Two other important parameters of cardiac function obtained by analysis of the indicator dilution curve obtained across the central circulation (9) were also included. These were the pulmonary dispersive time (*td*), which is largely a measure of the mean transit time across the pulmonary vascular bed and from which a measure of pulmonary blood volume (DV/m^2) can be obtained ($td \times CI = DV/m^2$), and the time constant of washout of dye containing blood from the left heart (*tm*) from which a measure of the mean cardiac ejection fraction (EFx) and the left ventricular end-diastolic volume (LVEDV) during the period of CI determination can be computed (10).

From these primary data a frame of reference defined as the state of mean normal compensation (R state) for a group of older high-risk patients was derived. Each individual patient can be compared to this normal reference group by virtue of the standard deviation of each of the physiologic variables from the normal R state mean (zero standard deviation). By using a statistical clustering technique that grouped together individuals of similar physiologic patterns, these workers (6–8) showed that the entire spectrum of clinical severity in patients with various forms of trauma, sepsis, or cardiogenic critical illness could be viewed in terms of four pathophysiologic states (A, B, C, and D). The prototypical or mean patient pattern for each of these abnormal states can be described by the physiological patterns of the multivariable means shown in one of the four patterns in the corners of Figure 1. More important, any individual patient at a given time in his clinical course can be compared to each of the mean states (R, A, B, C, D). Since all physiologic variables have been converted to a common dimension (the standard deviation from normal), the patient's similarity to each state can be quantitatively expressed as a "distance" from each of the prototype states. Thus the nature and degree of that patient's physiologic compensation can be quantified by his absolute distance from a given state or from his distance ratios (D/A, C/B) between states (11). The pattern of changing state distances over time provides a quantitative trajectory of recovery by which clinical courses can be evaluated and compared with regard to the adequacy of their physiologic compensation over time (2).

In physiologic terms the A state is a normal stress response seen in compensated sepsis and after trauma or a major operation. It is characterized by increased heart rate and cardiac index, improved contractility (reduced *tm* and increased EFx), and increased oxygen consumption index ($O_2CI = AVo_2Dif \times CI$). This sympathetic response adaption occurs without any evidence of metabolic or respiratory abnormality. The failure to achieve an A state response in the presence of a major stress is an abnormal response pattern. The B state is also a hyperdynamic state, but it represents a more severe stage of deterioration in the septic process. The increased CI and other evidences of sympathetic response (↑EFx, ↑HR, ↓ET) are attempts to supply peripheral needs. In spite of this, oxygen consumption is reduced because of a narrowed AVo_2Dif, which apparently reflects a cellular O_2 extraction failure; a metabolic acidosis always occurs, though it may be masked by a compensatory respiratory alkalosis. In the C state, respiratory decompensation is superimposed on the unbalanced septic process seen


PATIENT: 44 YEAR FEMALE WITH PANCREATITIS
INFECTED PANCREATIC PSEUDOCYST
POST OP CYSTOGASTROSTOMY
A STATE
B STATE
POST OP DAY 5
POST OP DAY 6
10:00 AM
HYPERDYNAMIC
CI 6.37
HR 110.00
MAP 105.00
VpH 7.48
VpCO2 35.00
Tm 3.40
Td 2.80
ET 0.29
RAP -2.00
AV DIF 3.50
VpO2 38.00
-2
0
+2
+4
D/A = 2.12
C/B = 1.65
NO DRUGS
9:22 AM
EARLY CARDIOGENIC
CI 2.75
HR 130.00
MAP 73.00
VpH 7.49
VpCO2 44.00
Tm 7.23
Td 4.20
ET 0.20
RAP 6.50
AV DIF 4.26
VpO2 37.00
-2
0
+2
+4
D/A = 1.26
C/B = 1.37
DIGOXIN 0.250 mg
13:31 PM
HYPERDYNAMIC
CI 5.05
HR 120.00
MAP 90.00
VpH 7.50
VpCO2 43.00
Tm 3.96
Td 2.52
ET 0.25
RAP 3.50
AV DIF 2.63
VpO2 43.00
-2
0
+2
+4
D/A = 2.60
C/B = 2.09
DIGOXIN = 0.125 mg
ISUPREL 0.5 μg/min
D STATE
C STATE

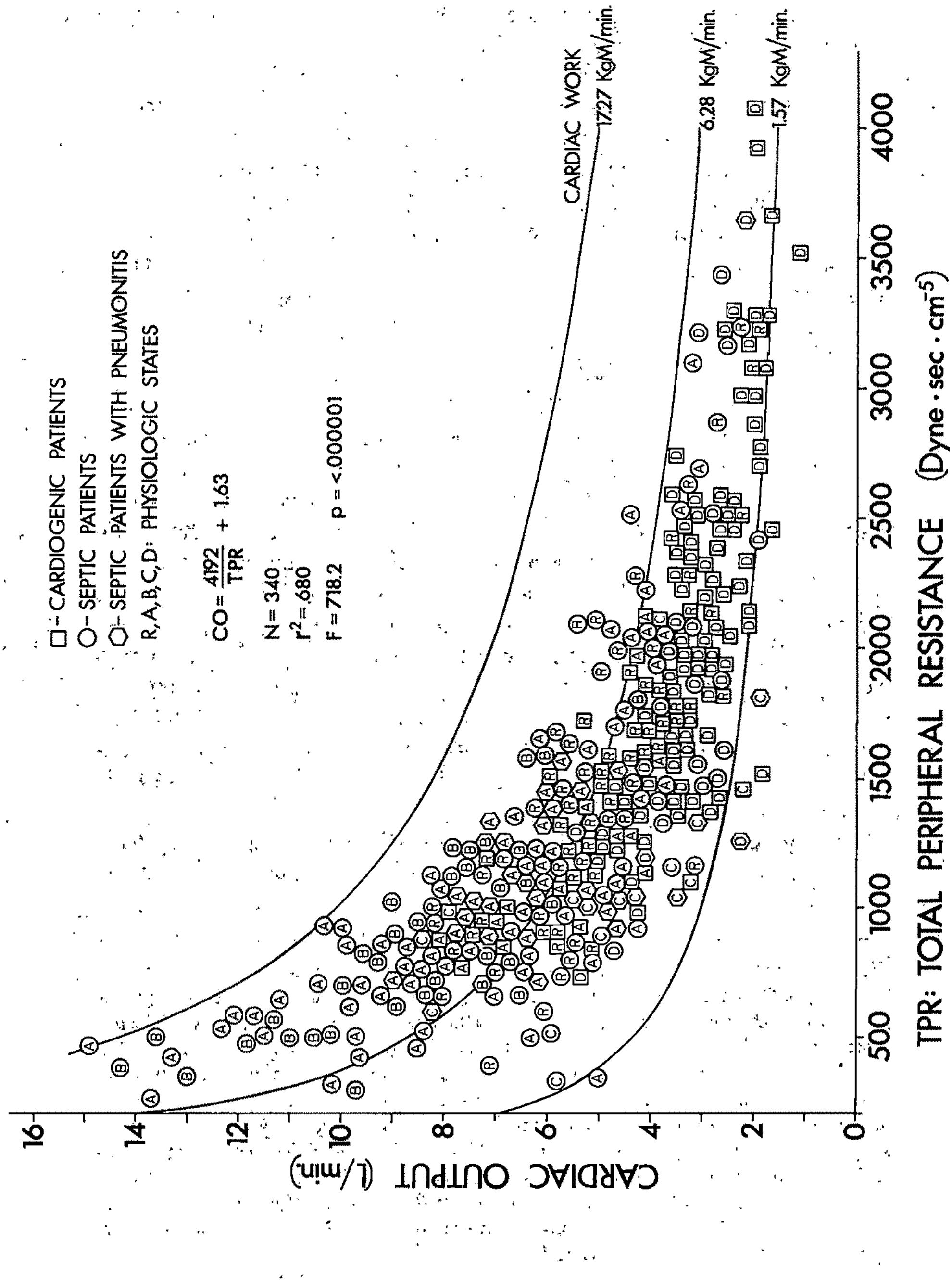
in the B state, and profound respiratory and metabolic acidosis occurs. In contrast to the B and C states, the D, or cardiogenic, state represents a pattern of primary myocardial rather than peripheral failure. In this state, there is a decrease in myocardial contractile function manifested by a prolonged *tm* and a reduced EFx. The cardiac index falls and the $AV_{O_2}Dif$ widens as oxygen extraction increases to compensate for the low flow state. Unless hypovolemia is also associated with the cardiac failure as may occur in postsurgical hemorrhage, the pulmonary blood volume (DV/m^2) rises. Hypotension and acidosis may occur. While this pattern is characteristic of acute myocardial infarction and may occur in post cardiac surgical patients, it can also occur as a result of septic myocardial decompensation (Figure 1), in association with or following a period of profound B state hyperdynamic metabolic insufficiency (2, 10).

Cardiovascular and Cardiopulmonary Relations

The classic cardiovascular relationships in sepsis can be interpreted from the perspective of the adequacy of the hemodynamic response to the metabolic needs. Figure 2 shows the relationship between the peripheral vascular resistance (TPR) and the cardiac output (CO) for septic and nonseptic patients labelled by their nearest physiologic state distance. This figure also shows lines of constant cardiac minute work. Normal older patients have cardiac output levels of 5–6 liters/min at a TPR range of 1200–1500 dyne • sec • cm^{-5} and do a mean cardiac work of about 6 kg • m/min. Hyperdynamic septic patients in A, and especially in B, states have cardiac outputs that are generally greater than 8 liters/min and a TPR level of < 1100. This level of vascular pressure:flow relationship (vascular tone) is disproportionately lower than normal and appears to represent the primary physiologic vascular lesion in sepsis (12).

The high cardiac output maintained by septic patients appears to have a distinct survival value (3). With this in mind, note that in spite of the reduced afterload hyperdynamic septic patients and especially the B state

Figure 1 Circle diagrams of physiologic states. At corners are shown the prototype patterns for the A, B, C, and D states. The three large circles in the center are the physiological septic patterns manifested by a 44-year-old woman with an infected pancreatic pseudocyst. The real values of physiological variables are given together with variable name on each day. The perfect circle in the center of each pattern equals zero standard deviations from reference control R state. Each dotted line represents one standard deviation from R either increased or decreased. Below each patient pattern are the state distance ratios D/A and C/B from the state prototypes. The physiologic state classification of this septic patient was A state during initial phase of the septic acute respiratory distress syndrome (10:00 am Post Op Day 5), early myocardial depression (D state) (9:22 am Post Op Day 6), and B state after inotropic resuscitation (13:31 pm Post Op Day 6) showing that myocardial depression was associated with A to B state transition. (From Reference 10.)

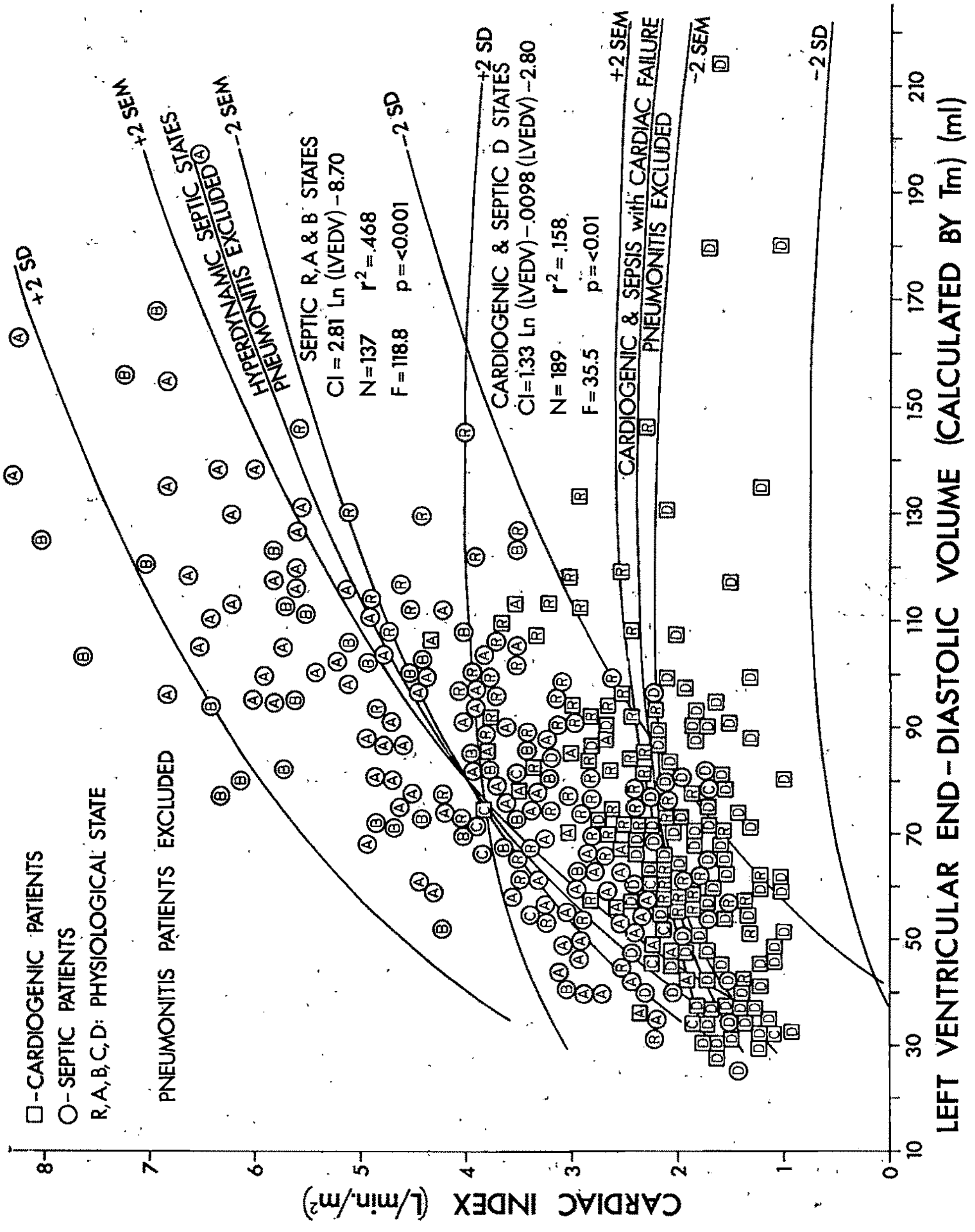

□ - CARDIOGENIC PATIENTS
○ - SEPTIC PATIENTS
⬡ - SEPTIC PATIENTS WITH PNEUMONITIS
R, A, B, C, D: PHYSIOLOGIC STATES
CO = 4192/TPR + 1.63
N = 340
r² = .680
F = 718.2
p = <.000001
CARDIAC WORK
17.27 KgM/min.
6.28 KgM/min.
1.57 KgM/min.
CARDIAC OUTPUT (L/min.)
16
14
12
10
8
6
4
2
0
500
1000
1500
2000
2500
3000
3500
4000
TPR: TOTAL PERIPHERAL RESISTANCE (Dyne · sec · cm⁻⁵)

unbalanced metabolic responders are forced to perform very large amounts of cardiac work, ranging from 10–17 kgM/min (Figure 2)(10), similar to that seen in heavy exercise (13). This increased level of cardiac work is maintained for days or even weeks as the septic process progresses or resolves. This requires very large substrate fuel utilization and energy production. In the hyperdynamic B state septic patients, in whom oxidative metabolism appears to be impaired (2, 7), the cardiac work may explain nearly 50% of the variability in the total oxygen consumption. With continuing energy requirements of this magnitude, not surprisingly the incidence of occult and overt myocardial depression in sepsis is relatively high and D state cardiogenic decompensation is common (Figure 2).

Cardiogenic decompensation in sepsis is associated with a higher incidence of death (7, 12). Moreover, as shown in Figure 3, the Starling-Sarnoff relationships are poorer for D state septics (where they are similar to the relations seen in nonseptic cardiogenic shock patients) compared to A and B state hyperdynamic septics. Consequently, the use of cardiac inotropic support (3, 4, 14) and sometimes vasodilator (15) agents is critically important to survival in these patients. Figure 3 also makes another important point. The B state patients with very low vascular resistance and a reduced O_2CI are mainly the ones who have the most left-shifted cardiac index to LVEDV ratio, lying at or above the +95% confidence limit for the Starling-Sarnoff relations in hyperdynamic septic patients (2). This suggests that the great flow increase seen in B state patients is a sympathetically mediated compensation to the abnormal vascular tone relationship; this is confirmed by the fact that these patients generally have a very large EFx (10).

The peripheral cardiovascular adaptive response seen in the septic patient is mirrored in the cardiopulmonary ventilation:perfusion ($\dot{V}_A/\dot{Q}_T$) relationships. This is shown in Figure 4, where the mean $\dot{V}_A/\dot{Q}_T$ is shown as a function of CI in a group of critically ill patients maintained on continuous ventilatory support. The hyperdynamic septic A and B state patients, who have the highest levels of CI ($>$ 4 liters/min/m^2, where 3 is normal) also have the lowest levels of $\dot{V}_A/\dot{Q}_T$ ($<<$ 0.6, where 0.8 is normal). Indeed the patients with the lowest $\dot{V}_A/\dot{Q}_T$ are those in the B state of metabolic imbalance who manifest the lowest TPR (Figure 2) and who have the highest CI/LVEDV ratio (Figure 3) (2, 10). Since these patients also have the highest pulmonary shunt ($\dot{Q}_S/\dot{Q}_T$) and the greatest ventilatory dead space to tidal volume ratio (V_D/V_T), Siegel et al (10) suggested that the initial hypoxemia of sepsis results from the passive redistribution of pulmonary

Figure 2 Relation between cardiac output and total peripheral resistance labelled by physiologic state classification. Lines of constant cardiac work in kilograms per minute are shown. (From Reference 2.)


□-CARDIOGENIC PATIENTS
○-SEPTIC PATIENTS
R, A, B, C, D: PHYSIOLOGICAL STATE
PNEUMONITIS PATIENTS EXCLUDED
+2 SD
+2 SEM
HYPERDYNAMIC SEPTIC STATES Ⓐ
PNEUMONITIS EXCLUDED
-2 SEM
SEPTIC R, A & B STATES
CI = 2.81 Ln (LVEDV) - 8.70
N = 137
$r^2 = .468$
F = 118.8
p = <0.001
-2 SD
+2 SD
CARDIOGENIC & SEPTIC D STATES
CI = 1.33 Ln (LVEDV) - .0098 (LVEDV) - 2.80
N = 189
$r^2 = .158$
F = 35.5
p = <0.01
+2 SEM
CARDIOGENIC & SEPSIS with CARDIAC FAILURE
PNEUMONITIS EXCLUDED
-2 SEM
-2 SD
CARDIAC INDEX (L/min./m²)
0
1
2
3
4
5
6
7
8
LEFT VENTRICULAR END-DIASTOLIC VOLUME (CALCULATED BY Tm) (ml)
10
30
50
70
90
110
130
150
170
190
210

perfusion. This is attributed to the abnormal reduction in vascular tone in the presence of a heightened cardiac inotropic response (EFx↑). This reduces LVEDV, which in turn reduces LVEDP and thus LAP relative to the increase in flow, causing greater perfusion of dependent lung segments and underperfusion of the superior lung segments. This is consistent with the fact that the early hypoxemia and increased $\dot{Q}_S/\dot{Q}_T$ seen in sepsis occur at a time when the lung x ray is clear (10, 16, 17). Only later when septic permeability changes (18, 19) does x-ray evidence of microatelectasis and consolidation appear in the dependent lung (the posterior segments since these patients are generally supine in bed). The therapeutic implications of these interrelationships are very great; evidence of hyperdynamic sepsis, especially during the A to B state septic transition, mandates careful attention to ventilatory support (10). Indeed, early use of intubation and volume ventilation with end-expiratory pressure may be necessary to increase the markedly reduced $\dot{V}_A$ (Figure 4) and force ventilation into underventilated, overperfused lung segments. Also, in the markedly hyperdynamic patient with a high EFx and low LVEDV, cautious infusion of colloid volume to expand intravascular volume can raise LVEDV and thus redistribute pulmonary perfusion by increasing the $\dot{V}_A/\dot{Q}_T$ (10, 20). The initial pulmonary insufficiency of sepsis is thus a consequence of the sympathetic compensation for the metabolically induced alterations in vascular tone and is also a reflection of the physiologic state.

Physiologic and Metabolic Correlations

Evidence that the physiologic state is a manifestation of fundamental metabolic abnormalities has been presented by Siegel et al (2, 11) and others (21, 22). In particular physiologic patterns reflect a more fundamental set of abnormalities in glucose, amino acid, and fat metabolism (2). Figure 5 shows the patterns of abnormalities in the circulating blood of glucose metabolites, branched-chain amino acids (BCAA), aromatic amino acids (AAA), amino acids involved in urea synthesis, fat metabolites and ketone bodies, and glucose-regulating hormones, as they relate to the abnormalities of cardiovascular function and the state distance ratios D/A and C/B. The means of sets of observations on three groups of patients (traumatic nonseptic, septic A state, and septic B state) are compared with regard to the magnitude of their standard deviational differences (SD) from the means of

Figure 3 Relation between cardiac index and left ventricular end-diastolic volume computed from stroke volume (SV) and the *tm* cardiac ejection fraction (EFx) (LVEDV = SV/EFx). Also shown are the mean, the standard error of the mean, and two-standard-deviational spread of each of the regression equations for the hyperdynamic (R, A, B) septic patients and the cardiogenic nonseptic and septic D state patients. The difference between slopes of the regressions has significant $P < 0.001$. (From Reference 2.)

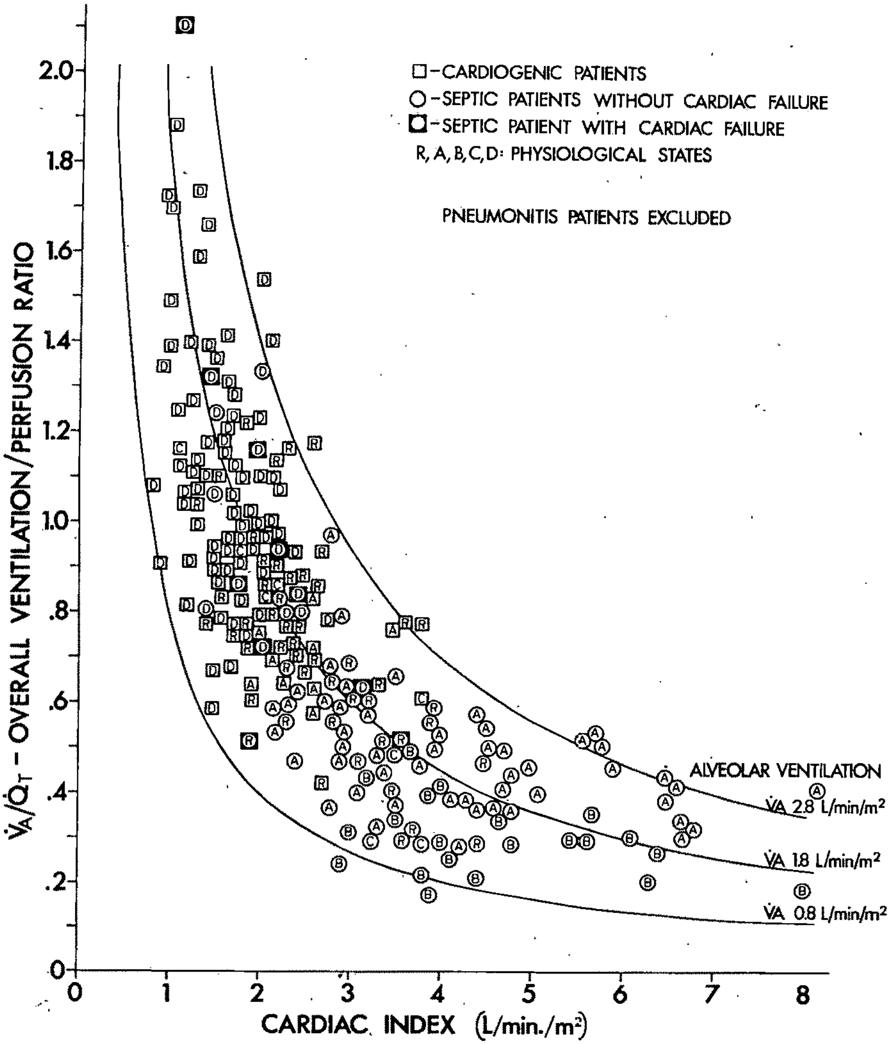


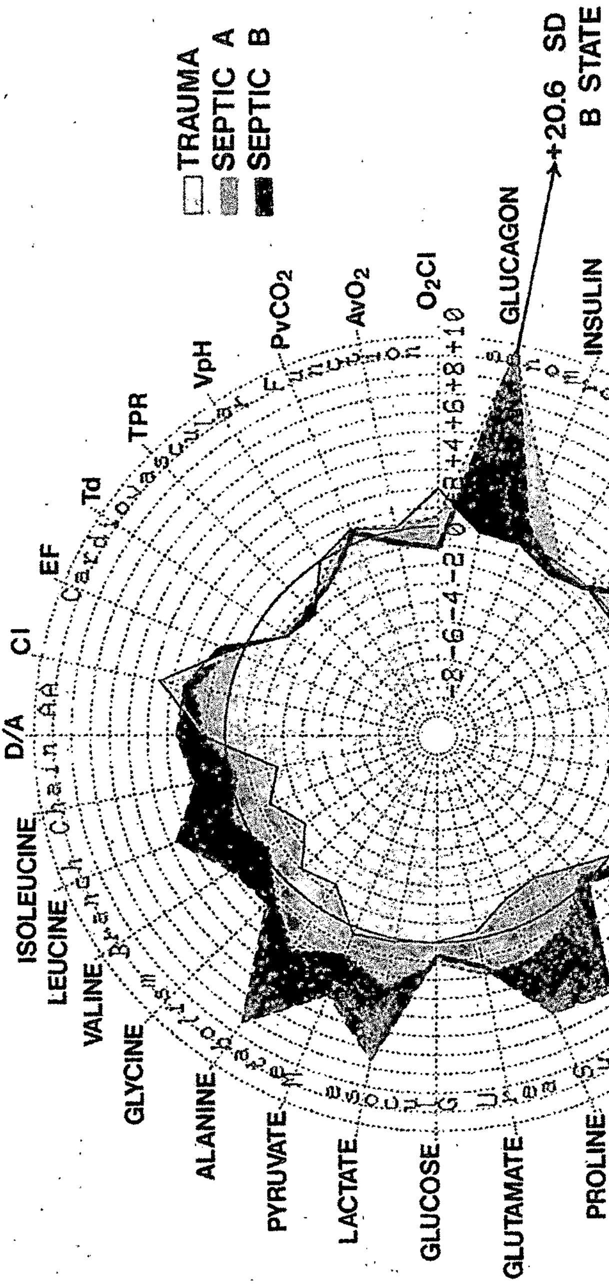
Figure 4 Relation between overall ventilation: perfusion ratio ($\dot{V}_A/\dot{Q}_T$) and cardiac index labelled by physiological state. Also shown are lines of constant alveolar ventilation. All patients were being maintained on constant ventilator support at minute volumes of 9–24 liters/min. (From Reference 2.)

a group of control elective general surgical patients in the R state of accommodation (the perfect circle at zero standard deviations). A physiologically important alteration seems likely for a mean change of ±1 SD from control. The nonseptic traumatized patients exhibit a pattern of increased CI and O_2CI associated with reduced circulating levels of BCAA, AAA, and the precursors of urea, with normal or slightly elevated levels of urea and glucose. There are also normal triglycerides, free fatty acids (FFA), and ketone levels with normal levels of glucose-regulating hormones glucagon and insulin.

In contrast, as sepsis progresses from compensated A state to the unbalanced B state, in which the progressively more abnormal pressure:flow relations are associated with a fall in oxygen consumption and a metabolic acidosis, one sees a progressive rise in BCAA with a marked increase in the levels of the NH_3^+ carrier amino acids alanine and glycine (2, 11, 22). Pyruvate and lactate rise proportionately (2), as does the level of circulating glucose. All of the amino acids and intermediates of urea synthesis (proline, ornithine and arginine, etc) rise as does urea, even without any evidence of renal failure (23). However, the most marked abnormalities in amino acid metabolism are seen in the increases in aromatic amino acids phenylalanine and tyrosine (2, 24). Under these circumstances A. Nespoli et al (unpublished results) observed an increase in octopamine and phenylethamine. The levels of sulfur-containing amino acid, methionine, and its metabolic catabolic product α-aminobutyric acid also rise (24). Triglycerides rise to extremely high levels even though FFA levels remain similar to those seen in trauma; there is a progressive elevation in the ratio of reduced to oxidized ketone bodies due to a rise in β-hydroxybutyric acid with normal or low levels of acetoacetate (2). The most striking changes from the nonseptic traumatic injured patient lie in the distorted pattern of insulin and glucagon; glucagon levels rise to more than 20 SD above normal in B state sepsis, with lesser rises in insulin so that the glucagon to insulin ratio progressively increases (2, 24). The mean value for the circulating levels of these hormones and substrates are shown in Figure 6, which also displays the metabolic pathways for oxidative catabolism in an idealized hepatocyte.

Septic Autocannibalism

As B state progression occurs, so does the phenomenon of "septic autocannibalism" (22), which is highly correlated with a fatal outcome. In septic autocannibalism, the normal dependence of the circulating levels of many amino acids, and especially the BCAA and the NH_3^+ carrier amino acids alanine and glycine, on the quantity of exogenously administered amino acids is lost. In these mostly B state patients (or A state patients close to a B state transformation by state distance ratios), more BCAA and isoleucine appear over and above that which can be accounted for by the administered nutrition (22). These high "endogenous" BCAA levels occur simultaneously with increased levels of AAA, proline, and methionine and are correlated with increases in alanine and glycine (22). This suggests enhanced muscle catabolic release of protein-derived amino acids at a time of increased utilization of BCAA. Such high "endogenous" BCAA levels


☐ TRAUMA
▨ SEPTIC A
■ SEPTIC B
+20.6 SD
B STATE
-8 -6 -4 -2 0 +2 +4 +6 +8 +10
Cardiovascular Function
Hormones
Fat Metabolism
Aromatic AA
Urea Synthesis
Gluconeogenesis
Branch Chain AA
O2CI
GLUCAGON
INSULIN
GLYCEROL
AcAc
BOHB
FFA
TRIG
C/B
TYROSINE
PHENYLALA.
UREA
ARGININE
ORNITHINE
PROLINE
GLUTAMATE
GLUCOSE
LACTATE
PYRUVATE
ALANINE
GLYCINE
VALINE
LEUCINE
ISOLEUCINE
D/A
CI
EF
Td
TPR
VpH
PvCO2
AvO2

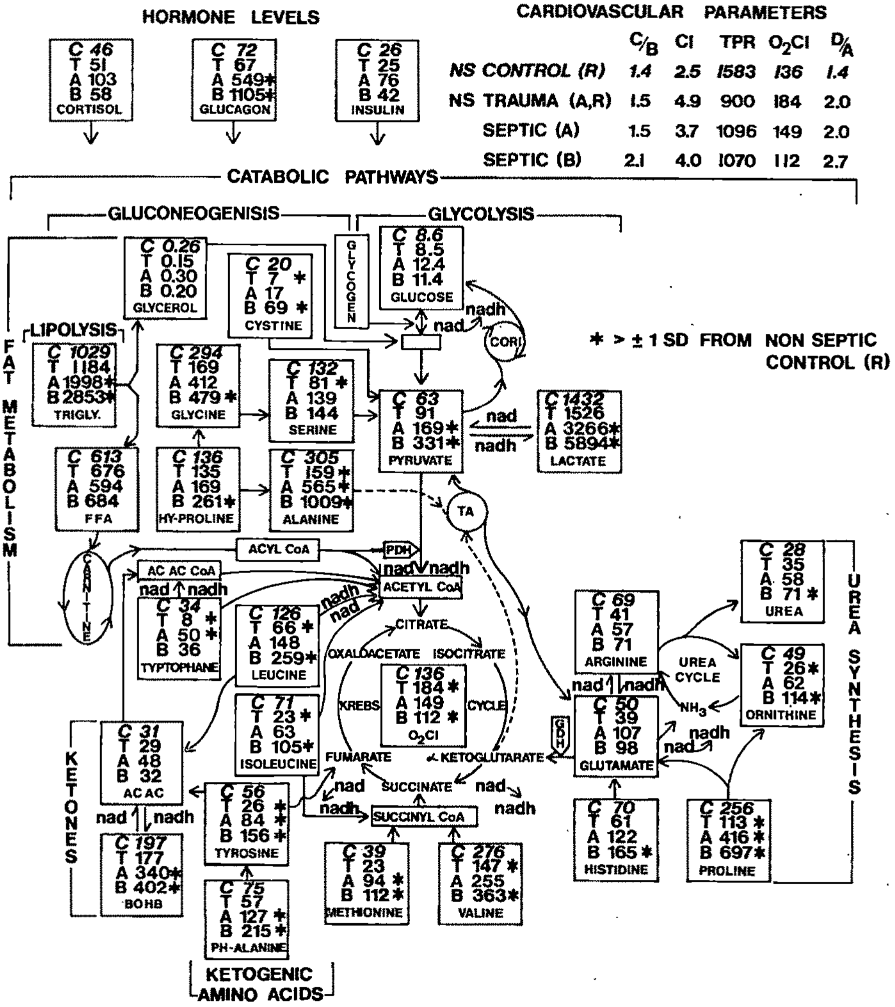
were not prevented by very large quantities of administered calories and amino acids. Moreover, there appears to be differential utilization of BCAA under these conditions, with impaired valine and isoleucine utilization but increased utilization of leucine (22). This clinical observation is compatible with the leucine utilization studies reported by Clowes et al (25) in isolated skeletal muscle from septic animals and patients. Differential utilization of BCAA under conditions of B state deterioration in which metabolic acidosis occurs may reflect preference for one pathway of BCAA access into the Krebs cycle over another, perhaps because of limited availability of a cofactor such as oxidized NAD^+, which plays a critical role in the BCAA ketoacid oxidation pathways (26).

Septic Multiple Organ Failure

The patterns of nonutilization seen in the B state septic process suggest a major pathologic alteration in the normal interorgan regulation of substrate and fuel-energy metabolism. Figure 7 outlines a suggested hypothesis for the failure mechanisms altering the normal muscle-liver-adipose tissue cycle of organ metabolism. In this conceptual model (2), the septic insult initiates an as yet unknown factor, which may be kinin related (27), and causes protein catabolism in skeletal muscle (21, 28–30). Neuroendocrine response is apparently enhanced with increased release of glucagon, corticosteroids, and catecholamines (2, 31). The muscle catabolism releases muscle amino acids in quantities proportionate to their concentrations in actin and myosin (25, 29, 30). However, the muscle itself can use only BCAA effectively as metabolic substrates (21, 32); as sepsis progresses into B state only leucine may be fully utilized (22). These BCAA are converted to their respective ketoacids, which then enter the Krebs tricarboxcylic acid (TCA) cycle for oxidation (Figure 6) (26). The NH_3^+ liberated by ketoacid formation in muscle is detoxified by transamination with pyruvate, derived from muscle glucose-6-PO_4 or glycogen, to form the carrier amino acids alanine and glycine (33). These carrier amino acids, which are markedly increased in sepsis (2, 22), are transported via the blood to liver and kidney where the transamination is reversed and alanine plus α-ketoglutarate yields pyruvate and glutamate (34).

In the nonseptic patient the major part of the pyruvate produced by this transport mechanism appears to be oxidized via acetyl-CoA to the TCA

Figure 5 Physiological and metabolic circle diagram for patients with nonseptic major trauma, and patients in septic A and B states. The perfect circle at zero standard deviations represents the mean values for a control group of general surgical patients in the R state of physiological compensation. All other mean lines are normalized by the standard deviation of the control group. Data are grouped in terms of appropriate physiological-metabolic functional interrelations. (From Reference 2.)



CARDIOVASCULAR PARAMETERS

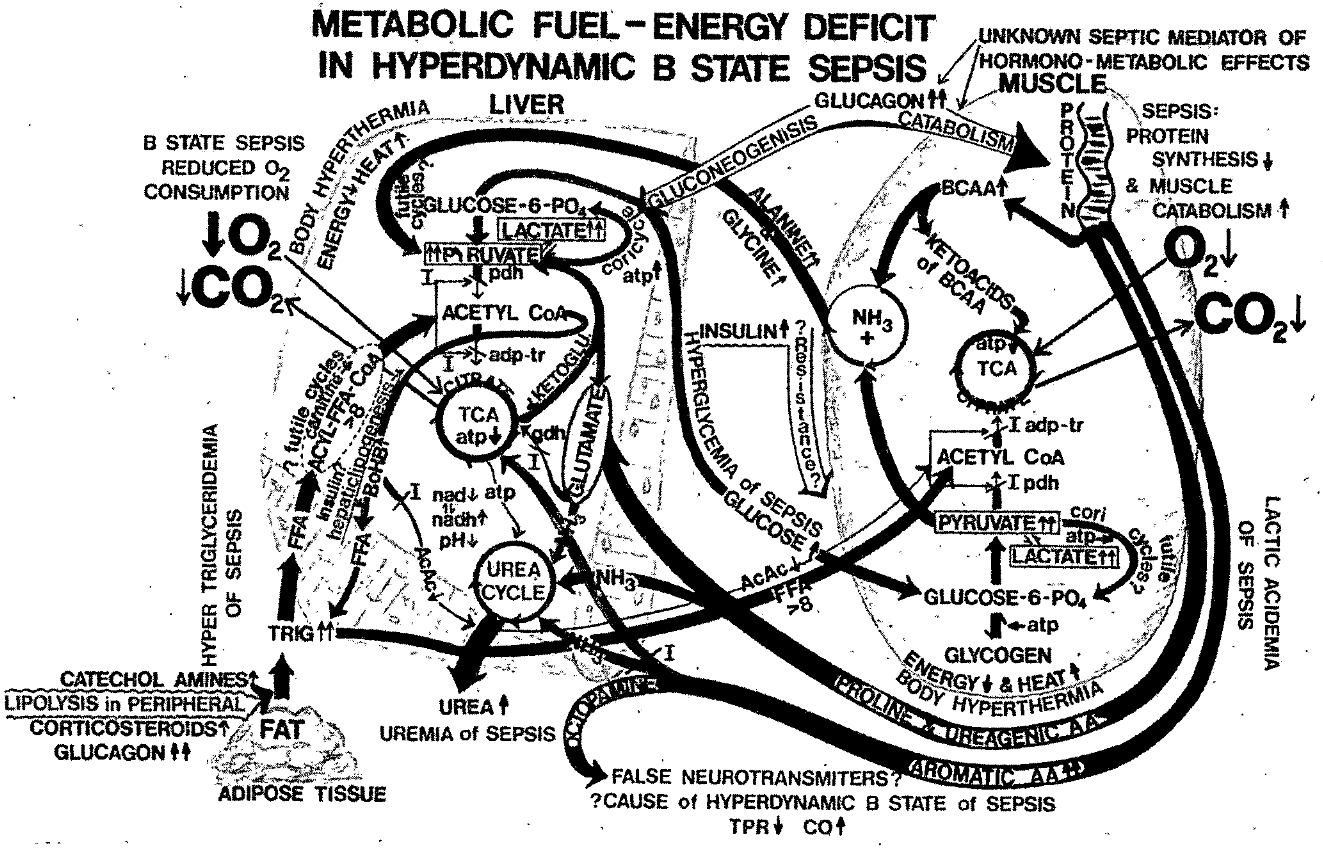
	C/B	CI	TPR	O_2CI	D/A
NS CONTROL (R)	*1.4*	*2.5*	*1583*	*136*	*1.4*
NS TRAUMA (A,R)	1.5	4.9	900	184	2.0
SEPTIC (A)	1.5	3.7	1096	149	2.0
SEPTIC (B)	2.1	4.0	1070	112	2.7

Figure 6 Metabolic map of known catabolic pathways in liver for amino acids, fat metabolites, glucose precursors, and precursors for urea synthesis. Values indicate means of physiological study groups. (*C* equals nonseptic control general surgical patients in R state; *T* equals nonseptic traumatized patients in A or R state; *A* equals septic A state patients; *B* equals septic B state patients. Cardiovascular parameters for these groups are shown in the right upper corner.) Amino acids μmoles/liter, glucose mM/liter, glucagon pg/ml, insulin μU/ml, O_2CI ml/min/m². (From Reference 2.)

cycle, and oxygen consumption rises. However, in the septic patient, especially as B state sepsis progresses (Figures 5, 6), there is evidence of diversion of pyruvate into gluconeogenic pathways (Figure 7), with a higher level of equilibration with lactate (2)(Figure 6). In the septic liver, a significant fraction of the glucose-6-PO_4 produced by gluconeogenesis is apparently hydrolyzed producing glucose (35), which is then transported via the circulating blood to the visceral organs and skeletal muscle. This increased gluconeogenesis may account for the hyperglycemia seen in sepsis (Figure 7).

The nongluconeogenic amino acids liberated by muscle catabolism are also transported to the liver and kidney for oxidative catabolism and urea synthesis (34). However, as B state sepsis develops there appears to be impairment in the complete oxidation of the group of amino acids entering into the TCA cycle via glutamate to α-ketoglutarate, perhaps because of inhibition of glutamate dehydrogenase (GDH). Consequently, proline and the related ureagenic precursor amino acids increase (23), even though urea synthesis is also increased—this accounts for the uremia of sepsis. In the specific case of the AAA tyrosine and phenylalanine the impairment to oxidative catabolism is very great. Not only do the circulating levels of these substances rise (2, 24, 30), but there is evidence that false neurotransmitter byproducts, such as octopamine (Nespoli, unpublished observations) also rise and this may account for the close relationship of the abnormal vascular tone and $\dot{V}_A/\dot{Q}_T$ abnormalities to the rise in AAA (2).

The abnormalities of function in the third major organ of substrate fuel production, adipose tissue (Figure 7), appear to be correlated with the pathophysiologically increased levels of glucagon and may also be modulated by the rise in corticosteroids and catecholamines. The net result is increased peripheral lipolysis and the hypertriglyceridemia of sepsis. Since FFA and glycerol levels are not increased above those seen in nonseptic trauma, increased triglyceride breakdown into these two byproducts is likely occurring, with enhanced utilization but incomplete oxidation since O_2CI falls. Glycerol seems to be a significant gluconeogenic precursor in the septic patient (36) and FFA, especially the carnitine-dependent long-chain components, may be converted to the acyl-FFA-CoA form. A distortion of the normal mechanisms may explain the peculiar manifestations of the B state septic process in which oxidative metabolism decreases as various catabolic breakdown products and energetic fuel substrates increase in the plasma. The hepatic aspect could occur by virtue of the known inhibitory and stimulatory actions of long-chain FFA acting through the medium of acyl-FFA-CoA esters (37) in the presence of the increased gluconeogenic precursors alanine, pyruvate, lactate, and glycerol. These may be modulated by the peripheral protein catabolic, gluconeogenic and lipolytic actions of the elevated glucagon levels (38–40). The inhibitory action of these acyl-FFA-CoA esters on hepatic pyruvate dehydrogenase (PDH)(37), which would reduce pyruvate entry into the TCA cycle, and the concomitant stimulation of pyruvate carboxylase activity (41), which initiates gluconeogenesis, would make glucose synthesis dependent on the concentration of pyruvate, which is increased as B state sepsis develops (2)(Figure 6). The


METABOLIC FUEL-ENERGY DEFICIT
IN HYPERDYNAMIC B STATE SEPSIS
UNKNOWN SEPTIC MEDIATOR OF
HORMONO-METABOLIC EFFECTS
MUSCLE
LIVER
GLUCAGON ↑↑
CATABOLISM
GLUCONEOGENISIS
PROTEIN
SEPSIS:
PROTEIN
SYNTHESIS ↓
& MUSCLE
CATABOLISM ↑
B STATE SEPSIS
REDUCED O_2
CONSUMPTION
↓O_2
↓CO_2
O_2↓
CO_2↓
BODY HYPERTHERMIA
ENERGY↓ HEAT↑
futile cycles?
GLUCOSE-6-PO4
LACTATE↑↑
↑↑PYRUVATE
pdh
coricycle
atp↑
ACETYL CoA
adp-tr
CITRATE
KETOGLU
TCA
atp↓
gdh
GLUTAMATE
ALANINE↑↑
&
GLYCINE↑
BCAA↑
KETOACIDS
of BCAA
NH_3
+
atp
TCA
I adp-tr
ACETYL CoA
I pdh
PYRUVATE↑↑
cori
atp
LACTATE↑↑
futile cycles?
GLUCOSE-6-PO4
atp
GLYCOGEN
ENERGY↓ & HEAT↑
BODY HYPERTHERMIA
INSULIN↑
?Resistance?
HYPERGLYCEMIA of SEPSIS
GLUCOSE↑
AcAc↓
FFA
>8
? futile cycles
carnitine
ACYL-FFA-CoA
>8
insulin?
hepaticlipogenesis↓
BoHB↑
FFA
FFA
I
nad↓ atp
nadh↑
pH↓
AcAc
UREA
CYCLE
NH_3
NH_3
I
TRIG↑↑
HYPER TRIGLYCERIDEMIA
OF SEPSIS
CATECHOL AMINES↑
LIPOLYSIS in PERIPHERAL
CORTICOSTEROIDS↑
GLUCAGON ↑↑
FAT
ADIPOSE TISSUE
UREA ↑
UREMIA of SEPSIS
OCTOPAMINE
FALSE NEUROTRANSMITERS?
?CAUSE of HYPERDYNAMIC B STATE of SEPSIS
TPR↓ CO↑
PROLINE & UREAGENIC AA
AROMATIC AA↑↑
LACTIC ACIDEMIA
OF SEPSIS

increased glucagon levels may also be responsible for the enhanced gluconeogenesis by stimulating the second key gluconeogenic step, the conversion of pyruvate and oxaloacetate to phosphoenolpyruvate (40). This enhances the gluconeogenesis from alanine (39), which is also increased in sepsis by virtue of enhanced BCAA utilization in peripheral muscle. In addition, these acyl-FFA-CoA esters inhibit adenine nucleotide translocase (ADP-tr) in mitochondria and act as regulators for conversion of acetyl-CoA into AcAc, which is then reduced to β-hydroxybutyrate (BOHB) rather than into citrate formation and oxidation via the TCA cycle (37).

Other important effects of long-chain acyl-FFA-CoA esters are the inhibition of the malate-aspartate shuttle and the related transport of H^+ ions across the mitochondrial membrane resulting in a decreased mitochondrial redox potential (37). Also, the increased level of lactate in the presence of lactate dehydrogenase may influence H^+ availability in the cytoplasmic portion of the cell (42). This would produce intracellular and intramitochondrial acidosis and a tendency to saturate the nicotinamide adenine dinucleotide hydrogen acceptor, so that NADH rather than NAD^+ would occur in preponderance (42). The lack of a hydrogen acceptor in the presence of a reduced redox potential may be responsible for the lack of oxidation of BOHB to acetoacetate (AcAc) seen in the septic process, and may account for the rising BOHB/AcAc ratio (2). In addition, many of the critical enzymes in amino acid catabolism depend on NAD $\rightleftharpoons$ NADH or NADP $\rightleftharpoons$ NADPH (26), especially glutamate dehydrogenase (GDH) as well as key enzyme systems for the oxidation of leucine, isoleucine, valine, and tryptophane. This may well account for the generalized impairment in amino acid metabolism that occurs late in the B state, where all amino acid levels, including BCAA, rise as a preterminal event (2, 22). All of these factors would tend to reduce the amount of substrate entering the tricarboxcylic acid (TCA) cycle. The expected fall in oxygen consumption and CO_2 production appear to occur as the A to B state septic transition develops (2).

An additional critical consideration in the septic process is this: when oxidative metabolism and therefore ATP synthesis are declining, two processes that are extremely competitive for hepatic ATP, namely, urea synthesis and gluconeogenesis (34), are increasing. At the reduced level of oxygen consumption in the septic B state patient, it is possible that a major fraction of the already diminished ATP produced is being diverted into these two synthetic cycles rather than supporting other catabolic metabolic pathways and protein synthesis, which are also energy dependent. Indeed, the synthesis of acute phase proteins and immunologic proteins decreases and skin test

Figure 7 Diagram of multiple organ fuel-energy deficit in B state sepsis.

hypersensitivity is reduced even to the point of complete anergy in B state sepsis (F. B. Cerra et al, unpublished observations).

Another possible consequence of the increased gluconeogenesis at a time when pyruvate access to oxidative metabolism through acetyl-CoA appears inhibited is the production of a futile cycle in which pyruvate is synthesized to fructose-6-PO_4 through the gluconeogenic pathway and then returned by glycolysis to pyruvate without oxidation (43). This process is extremely wasteful of ATP, produces increased heat and net reduction in energy, and may be related to the body hyperthermia characteristic of sepsis. Another possible futile cycle is the diversion of acetyl-CoA away from citrate into acetoacetate and then into β-hydroxybutyrate (37). Since β-hydroxybutyrate appears not to be oxidized to acetoacetate, hepatic lipogenesis may tend to increase with a net diversion of acetyl-CoA into resynthesis of FFA and triglycerides rather than into fat oxidation. This may account for the histologic fatty degeneration seen in the livers of nonsurviving patients with severe sepsis. This concept is also supported by clinical observations that when the B state septic patient dies there is evidence of marked muscle catabolism, with wasting of all skeletal muscle, at a time when peripheral adipose tissue fat deposits remain prominent and there is marked hepatic fat deposition.

Conclusion

None of these subcellular enzymatic inhibitory functions have been proven in septic patients by biochemical analysis of human liver and skeletal muscle or by tracer studies of labelled compounds. However, the likelihood that sepsis is a disease of disordered metabolic control seems high and this realization should stimulate more productive investigations into the substrate utilization and energy regulation mechanisms in these conditions. It also seems likely that the selective utilization of metabolic fuels whose metabolic pathways are less inhibited by sepsis may permit key energetic processes to continue thus facilitating protein synthesis rather than catabolism. The close correlation between the biochemical evidence of disordered metabolism and the cardiovascular and pulmonary physiologic dysfunction suggests that the two are intimately related, perhaps through the production of vasoactive false transmitters that induce the fall in vascular tone. The correlate is that in addition to standard antibiotic therapy and surgical drainage of septic processes, efforts at cardiac and respiratory support are critical to immediate survival of the septic patient, while the fundamental defects in intermediary metabolism and energy production are being treated by nutritional support measures that address the specific biochemical defects of septic multiple organ failure.

Literature Cited

1. Wiles, J., Cerra, F. B., Siegel, J. H., Border, J. R., McMenamy, R. H. 1980. The systemic response: does the organism matter: *J. Crit. Care Med.* 8:55–60
2. Siegel, J. H., Cerra, F. B., Coleman, B., Giovannini, I., Shetye, M., Border, J. R., McMenamy, R. H. 1979. Physiological and metabolic correlations in human sepsis. *Surgery* 86:163–93
3. Clowes, G. H. A., Vucinic, M., Weidner, M. G. 1966. Circulatory and metabolic alteration associated with survival or death in peritonitis. *Ann. Surg.* 163:866–85
4. MacLean, L. D., Mulligan, W. G., McLean, A. P. H., Duff, J. M. 1967. Patterns of septic shock in man—a detailed study of 56 patients. *Ann. Surg.* 166:543–62
5. Weil, M. H., Shubin, H., Biddle, M. 1964. Shock caused by gram-negative microorganisms: analysis of 169 cases. *Ann. Intern. Med.* 60:384–400
6. Siegel, J. H., Goldwyn, R. M., Friedman, H. P. 1971. Pattern and process in the evolution of human septic shock. *Surgery* 70:232–45
7. Siegel, J. H., Farrell, E. J., Miller, M., Goldwyn, R. M., Friedman, H. P. 1973. Cardiorespiratory interactions as determinants of survival and the need for respiratory support in human shock states. *J. Trauma* 13:602–19
8. Friedman, H. P., Goldwyn, R. M., Siegel, J. H. 1975. The use and interpretation of multivariate methods in the classification of stages of serious infectious disease processes in the critically ill. In *Perspectives in Biometrics,* ed. R. Elashoff, pp. 81–122. New York: Academic. 200 pp
9. Siegel, J. H., Farrell, E. J., Lewin I. 1972. Quantifying the need for cardiac support in human shock by a functional model of cardiopulmonary vascular dynamics: with special reference to myocardial infarction. *J. Surg. Res.* 13:166–81
10. Siegel, J. H., Giovannini, I., Coleman, B. 1979. Ventilation: perfusion maldistribution secondary to the hyperdynamic cardiovascular state as the major cause of increased pulmonary shunting in human sepsis. *J. Trauma* 19:432–60
11. Siegel, J. H., Cerra, F. B., Peters, D., Moody, E., Brown, D., McMenamy, R. H., Border, J. R. 1979. The physiologic recovery trajectory as the organizing principle for the quantification of hormonometabolic adaptation to surgical stress and severe sepsis. In *Advances in Shock Research,* ed. W. Schumer, J. J. Spritzer, B. E. Marshal, pp. 177–203. New York: Liss. 298 pp.
12. Siegel, J. H., Greenspan, M., Del Guercio, L. R. M. 1967. Abnormal vascular tone, defective oxygen transport and myocardial failure in human septic shock. *Ann. Surg.* 165:504–17
13. Epstein, S. E., Beiser, G. D., Stampfer, M., Robinson, B. F., Braunwald, E. 1967. Characterization of the circulatory response to maximal upright exercise in normal subjects and in patients with heart disease. *Circulation* 25: 1049–56
14. Siegel, J. H. 1976. Pattern and process in the evolution of and recovery from shock. In *The Aged and High Risk Surgical Patient: Medical, Surgical, and Anesthetic Management,* ed. J. H. Siegel, P. Chodoff, pp. 381–435. New York: Grune & Stratton. 920 pp.
15. Cerra, F. B., Hassett, J. M., Siegel, J. H. 1978. Use of vasodilator therapy in septic myocardial depression. *J. Surg. Res.* 25:180–83
16. Clowes, G. H. A. Jr., Hirsch, E., Williams, L., Kwasnik, E., O'Donnell, T. F., Cuevas, P., Saini, J. K., Moradi, I., Farizan, M., Saravis, C., Stone, M., Kuffler, J. 1975. Septic lung and shock lung in man. *Ann. Surg.* 181:681–92
17. Vito, L., Dennis, R. C., Weisel, R. D., Hechtman, H. B. 1974. Sepsis presenting as acute respiratory insufficiency. *Surg. Gynecol. Obstet.* 138:896–900
18. Staub, N. C. 1974. State of the art review. Pathogenesis of pulmonary edema. *Am. Rev. Resp. Dis.* 109:358–72
19. Sibbald, W. J. 1980. Pulmonary capillary permeability in cell injury. In *Shock, Anoxia and Ischemia, Pathophysiology, Prevention and Treatment,* ed. R. A. Cowley, B. F. Trump. Baltimore: Williams & Wilkins. In press
20. Krausz, M. M., Perel, A., Eimerl, D., Cotev, S. 1977. Cardiopulmonary effects of volume loading in patients in septic shock. *Ann. Surg.* 185:429–34
21. O'Donnell, T. F., Clowes, G. H. A., Blackburn, G. L., Ryan, N. T., Benotti, P. N., Miller, J. D. 1976. Proteolysis associated with deficit of peripheral energy fuel substrates in septic man. *Surgery* 80:192–200
22. Cerra, F. B., Siegel, J. H., Coleman, B., Border, J. R., McMenamy, R. H. 1980. Septic autocannibalism: a failure of exogenous nutritional support. *Ann. Surg.* 192:570–80

23. Cerra, F. B., Caprioli, J., Siegel, J. H., Border, J. R., McMenamy, R. H. 1979. Abnormal proline metabolism in human sepsis: evidence of specific metabolic blocks. *Ann. Surg.* 190:577–86
24. Cerra, F. B., Siegel, J. H., Border, J. R., McMenamy, R. H. 1979. The hepatic failure of sepsis: cellular vs. substrate. *Surgery* 86:409–22
25. Clowes, G. H. A., Randall, H. T., Cha, C.-J. M., Tyan, B. A., Hirsh, E. 1980. Amino acid and energy metabolism in septic and traumatized patients. *J. Parenteral Enteral Nutr.* In press
26. Rosenberg, L. E., Scriver, C. R. 1974. Disorders of amino acid metabolism. In *Duncan's Diseases of Metabolism,* ed. P. K. Bondy, L. E. Rosenberg, pp. 465–653. Philadelphia: Saunders. 948 pp.
27. Schumer, W., Erve, P. E., Miller, B. 1974. Immune response in septic shock. Therapeutic implications. In *Treatment of Shock: Principles and Practice,* ed. L. Nyhus, W. Schumer, pp. 141–53. Philadelphia: Lea & Febiger. 355 pp.
28. Border, J. R., Chenier, R., McMenamy, R. H., LaDuca, J., Seibel, R. W., Birkhan, R., Yu, L. 1976. Multiple systems organ failure: muscle fuel deficit with visceral protein malnutrition. *Surg. Clin. North Am.* 56:1147–67
29. Marchuk, J. B., Finley, R. J., Groves, A. C., Wolfe, L. I., Holliday, R. L., Duff, J. H. 1977. Catabolic hormones and substrate patterns in septic patients. *J. Surg. Res.* 23:177–82
30. Askanazi, J., Carpenter, Y. A., Michelsen, C. B., Elwyn, D. H., Furst, P., Gump, F., Kinney, J. M. 1980. Muscle and plasma amino acids following injury: influence of intercurrent infection. *Ann. Surg.* In press
31. Wilmore, D. W., Moyland, J. A. Jr., Lundsey, C. A., Falonna, G. R., Unger, R., Pruitt, B. A. Jr. 1973. Hyperglucagonemia following thermal injury: Insulin and glucagon in the posttraumatic catabolic state. *Surg. Forum* 24:99–101
32. Freund, H. R., Ryan, J. A. Jr., Fischer, J. E. 1978. Amino acid derrangements in patients with sepsis: treatment with branched chain amino acid rich effusions. *Ann. Surg.* 188:423–30
33. Felig, P. 1975. Amino acid metabolism in man. *Ann. Rev. Biochem.* 44:933–55
34. Krebs, H. A., Lund, P., Stubbs, M. 1976. Interrelations between gluconeogenesis and urea synthesis. In *Gluconeogenesis: Its Regulation in Mammalian Species,* ed. R. W. Hanson, M. A. Mehlman, p. 269, New York: Wiley. 592 pp.
35. Long, C. L., Kinney, J. M., Geiger, J. W. 1976. Nonsuppressability of gluconeogenesis by glucose in septic patients. *Metabolism* 25:193–201
36. Carpenter, G. A., Askanazi, J., Elwyn, D. H. 1979. Effects of hypercaloric glucose infusion on lipid metabolism in injury and sepsis. *J. Trauma* 19:649–54
37. Shrago, E., Shug, A., Elson, C. 1976. Regulations of cell metabolism by mitochondrial transport systems. See Ref. 34, pp. 221–38
38. Lefebvre, P. 1972. Glucagon and lipid metabolism. In *Glucagon: Molecular Physiology, Clinical and Therapeutic Implications,* ed. P. Lefebvre, R. H. Unger, pp. 109–19. New York: Pergamon. 370 pp.
39. Chiasson, J. L., Cook, J., Liljenquist, J. E., Lacy, W. W. 1974. Glucagon stimulation of gluconeogenesis from alanine in the intact dog. *Am. J. Physiol.* 227:19–23
40. Tilgham, S. M., Hanson, R. W., Ballard, F. S. 1976. Hormonal regulation of phosphoenopyruvate carboxykinase (GTP) in mammalian tissues. See Ref. 34, pp. 47–91
41. Barritt, J. G., Zander, G. L., Utter, M. F. 1976. The regulation of pyruvate carboxylase activity in gluconeogenic tissues. See Ref. 34, pp. 3–46
42. Williamson, J. R. 1976. Role of anion transport in the regulation of metabolism. See Ref. 34, pp. 165–220
43. Soling, J. Kleineke, J. 1976. Species dependent regulation of hepatic gluconeogenesis in higher animals. See Ref. 34, pp. 369–462

Ann. Rev. Med. 1981. 32:195–206
Copyright © 1981 by Annual Reviews Inc. All rights reserved

TREATMENT OF SICKLE CELL ANEMIA[1]

❖7391

Samuel Charache, M.D.

Departments of Medicine and Laboratory Medicine, Johns Hopkins Medical Institutions, Baltimore, Maryland 21205

INTRODUCTION

To date, controlled trials of treatment for sickle cell anemia have shown only that proposed therapies were ineffective. Treatments used by experienced clinicians have common elements, but no regimen could be called "outstanding" or even "successful." There are some very effective techniques that do work; they are common to all chronic diseases and include: (*a*) attention to complications that can be treated, (*b*) manipulation of the patient's environment, (*c*) "laying-on of hands," and (*d*) kindness. They are perhaps the most important modes of therapy in this disease, but only the first is discussed here.

TREATMENT OF VASO-OCCLUSIVE CRISES

Analgesia, Oxygen, and Intravenous Fluids

According to pharmacology texts (1, 2), 80–100 mg of meperidine given subcutaneously are equivalent to 10 mg of morphine, and the analgesia they produce lasts 2–4 hr (Table 1). The effects of 10 mg of morphine last 4–5 hr, as do those of oxycodone, while the analgesic effect of methadone lasts 3–5 hr. Morphine is poorly absorbed by mouth, meperidine is better, and codeine, oxycodone, and methadone are better still.

Oxygen is often administered to patients having painful crises, even though hyperbaric oxygenation has been shown to be ineffective in relieving pain. (3) There is little therapeutic advantage if the arterial Po_2 is raised

---

[1] Supported by Grant HL-02799 from the National Institutes of Health, and a grant from the Central Maryland Committee for Sickle Cell Anemia.

0066-4219/81/0401-0195$01.00

Table 1 Pharmacology of commonly used narcotics

Drug	Trade name	Parenteral dose (mg)	Duration of action (hr)	Efficacy of oral administration
Morphine	—	10	4–5	1/6–1/15
Hydromorphone	Dilaudid®	1.5	4–5	1/5
Oxycodone	Percodan®	10–15	4–5	1/2–1/3
Meperidine	Demerol®	80–100	2–4	1/4
Pentazocine	Talwin®	30–50	3–5	1/3–1/5
Propoxyphene	Darvon®	(180–240)	4?	—
Codeine	—	(120)		1/2–1/3
Methadone	Dolophine®	7.5–10	3–5	2/5

from 80 to 100 mm Hg, and administration of oxygen by face mask or nasal "prongs" can be quite uncomfortable. If the P_{O_2} is initially low (<75 mm Hg) because of shunting or perfusion of nonventilated lung, the situation is different because generalized sickling can occur. In this case, oxygen therapy must be vigorous, and the P_{O_2} must be rechecked to see what has been accomplished.

Most patients are given intravenous fluids to maintain hydration—even though they can and will drink fluids by mouth. Although more nursing care is needed to get patients to drink juice or carbonated beverages, preservation of veins for the future is effort well spent.

Addiction Hazards

Any narcotic can be addicting, but drugs like codeine and pentazocine (Talwin®) are perhaps safer than others because of the unpleasant side effects they produce if taken in large quantities. When instances of recognized narcotic abuse are tabulated, the addiction hazard of meperidine appears to be considerably less than that of propoxyphene (4). Since drug-seeking is learned behavior, the physician must control the dose and time of administration of narcotics (5) rather than give the patient the opportunity to seek the drug and have that behavior rewarded. There are some addicted sicklers; in my experience they all became addicted outside the hospital, and most admit that they are addicted. We give them a dose of meperidine (50–75 mg every 4 hours by injection) that would be effective in a nontolerant patient, but no more, and give them no narcotics to use at home.

Adjuvant Drugs

Hydroxyzine and phenergan are popular with our housestaff, for they are thought to potentiate narcotics. There is little evidence that they do so (6), but they are sedatives and they do keep patients quiet (although that is not the stated purpose). If patients are angry, agitated, or frightened, they should be sedated. Chlorpromazine is very effective in some patients, but so are soothing words if the physician can spare the time.

ACUTE COMPLICATIONS OF SICKLE CELL ANEMIA

Sickle Cell Lung Disease

This condition, characterized by fever, pleuritic pain, lung infiltrates, and often hypoxemia, is not uncommon in adults (7). Sputum cultures usually yield *Hemophilus* species, if any potential pathogen is found, and one is never sure whether he is treating infection or infarction. Nevertheless, if the patient is seriously ill, it would seem wise to administer ampicillin or a similar antibiotic.

Since hypoxemia produced by the lung lesion can lead to further sickling, it is important to monitor the arterial Po_2. If it is less than 75 mm Hg and does not rise into the normal range with oxygen therapy, partial exchange transfusion should be done promptly. Patients with advancing disease should have the Po_2 monitored very closely, and vigorous measures (intubation, positive end-expiratory pressure, etc) should be instituted if the situation deteriorates.

In children the situation is different, and fulminant pneumococcal infections are a real threat. Patients may appear only mildly ill, but die shortly thereafter if an appropriate antibiotic is not given. Prevention by means of polyvalent pneumococcal vaccine is moderately useful in children (8, 9); we have not given it to many of our adult patients since pneumococcal infections are uncommon in them, and some patients develop frightening fever and local reaction in response to vaccine injection (10). Prophylactic penicillin therapy carries with it the risk of resistant bacteria (11) but is used by some experienced clinicians.

Stroke

In children, stroke is usually due to occlusion of a large cerebral vessel, perhaps because vasa-vasora were occluded earlier (12). In adults, subarachnoid hemorrhage is more common (13). Partial exchange transfusion should be carried out before arteriograms are done (14), and oxygenation must be maintained. Children often survive, and probably should be maintained with chronic transfusion therapy for several years (15). Since recovery can occur (16), therapy should be aggressive.

Right Upper Quadrant Syndrome

An acutely ill patient with right upper quadrant pain, fever, and increased jaundice may have acute cholecystitis, extrahepatic biliary tract obstruc-

Table 2 Procedure for a partial exchange transfusion[a]

1. Use at least a #16 needle connected to a blood bag via an extension tube and a tube with a bayonet adapter[b]. Do not remove any blood until one unit of donor cells is on hand. Have 1000 cc of normal saline solution ready for infusion.
2. Remove 500 ml of blood. Inexperienced phlebotomists should ask the Blood Bank what a filled blood bag looks like, and a scale can be used (1 ml of blood weighs about 1.06 g). In *emergency* situations, blood can be "pulled" out and "pushed" in through a jugular or femoral venous catheter.
3. Infuse 500 ml of saline.
4. Remove a second 500 ml of blood. Save both units—they are valuable reagents for those studying sickle cells.
5. Infuse 5 units of packed red cells as fast as they will flow without prior dilution.
6. In extremely urgent situations, use a #14 needle, pull the blood out with a 50-ml syringe, and push in the normal cells (or use a pressure apparatus). Since rapid infusion of 1000 cc of red cells at 4°C produces severe shaking chills, pre-warm the blood.
7. See text if the starting hematocrit is >27%.
8. If a cell separator is available and the extra expense can be justified, elegant exchange transfusions can be performed with little risk.

[a] For a 50-kg adult with Hct 20–27%.
[b] AE-2—Fenwal Corporation.

tion, viral hepatitis, or one or more infarcts due to sickling in the liver (17, 18). Patients with chronic hemolytic anemia, and consequent increased formation of bilirubin, can develop spectacular degrees of jaundice when things go wrong in and around the liver; unfortunately, the bilirubin level per se does not differentiate between "surgical" and "nonsurgical" jaundice. Close medical and surgical co-operation is essential in managing these patients—some improve abruptly (by passing a stone?); in others, prompt surgery seems the least of several evils. Perhaps the most important thing for all to remember is that sickle cell infarcts and hepatitis can completely mimic biliary tract disease, and one must move cautiously.

One solution to the problem would be to remove asymptomatic gall stones as soon as they are detected. Many surgeons think that is a fine idea (19), and the risk of elective surgery is not great, but the risk from the stones may be less. I recommend cholecystectomy whenever symptoms become suggestive, but not before.

Partial Exchange Transfusion

It is probably good to do a partial exchange transfusion before general anesthesia, or whenever an acute situation is deteriorating. The technique outlined in Table 2 is designed for an adult of about 50 kg whose hematocrit is 20–27%. If the hematocrit is above 27%, prior removal of another 500 ml of blood (usually the night before the exchange) is recommended. In patients with Hb SC disease, with hematocrits well over 27%, removal of several units beforehand may be necessary; if the patient is kept in bed, and his blood volume replenished, a phlebotomy of 1000 cc is usually well tolerated. Whatever technique is used, the proportion of sickle cells (or Hb S) should be checked after transfusion, and additional normal red cells given if it is <50%.

"Female Troubles"

CONTRACEPTION There is no clear evidence whether the hazards of the "mini-pill" exceed the hazards of pregnancy. These pills may be safer and more efficacious than IUDs, but I prefer one of the latter if it can be tolerated. Both pills and IUDs are preferable to barrier and rhythm methods (which have been outstanding failures in my young patients).

PREGNANCY When a patient becomes pregnant, and wants the baby, she is followed closely in "High Risk" clinic. Although all pregnant sicklers are transfused prophylactically in some clinics (20–23), my criteria for transfusion are (*a*) increased severity of anemia with symptoms, or (*b*) a change in the severity or frequency of crises. The first situation is rare, since patients adapt remarkably well. The second situation does occur: a previously symptom-free patient gets severe bone pain, or a patient who has occasional bone pain gets "pneumonia."

When we transfuse, we aim for >50% normal cells and give additional two-unit transfusions approximately every four weeks until delivery, to maintain that level. If transfusion is for anemia alone, we usually administer only two units of packed cells.

Normal delivery should be done without inhalation anesthesia. If Caeserean section is required, spinal anesthesia would seem preferable, but venous stasis in the legs attendant upon paralysis of sympathetic nerves could present a problem. If inhalation anesthesia seems necessary after a dialogue with an experienced anesthesiologist, partial exchange transfusion should be done. In emergency situations, the transfusions can be done in a few hours; fetal monitoring during the procedure is probably important.

ABORTION The possibility of abortion is gently suggested to any sickler who becomes pregnant, but is not recommended unless the patient is dubious about continuing the pregnancy; the risks of pregnancy are high enough that only desired gestations should continue. All our young women are advised to consult us early if they think they're pregnant; that makes subsequent matters easier. During abortion, anesthesia or heavy sedation should be avoided. Using suction curettage in pregnancies of <14 weeks duration, we have encountered no problems with intravenous administration of meperidine. Because of the slight (but present) danger of respiratory arrest, I ask that an anesthesiologist be nearby until abortion is completed. We do not transfuse such patients; thus far, there have been remarkably few problems. In 16–22 week pregnancies, abortion is accomplished by injecting urea and prostaglandin F2 into the uterus. Usually analgesic requirements are not great; if pain relief is necessary, respiratory depression must be avoided.

CHRONIC COMPLICATIONS

Ankle Ulcers

Nearly complete bed rest remains the most effective, if least practical, method of treatment. Patients usually cannot stay in bed at home, and may require hospitalization—only to have the ulcer break down shortly after discharge. Some patients are helped by Unna boots; most do as well with frequent dressing changes. Wet-to-dry dressings with Burrow's solution provide gentle debridement, and may be better than more "surgical" removal of necrotic debris. I teach the patient (and sometimes the nurse!) how the dressings should be done, and then have the patient change them himself every two hours while awake whether at home or in the hospital. Topical antibiotics treat the doctor more than the patient; if there are signs of acute spreading infection (fever, lymphangitis, tender nodes) systemic drugs should be used.

Subacute Bone Disease

Most patients have pain in their bones during "vaso-occlusive" crises, and it is perhaps surprising that they do not have more orthopedic disability. Subacute problems range from the "hand-foot" syndrome seen in infants to joint effusions (24) and osteomyelitis. Bone marrow scans (25) have not been useful, particularly because one never knows how old the lesions are unless a prior scan was performed. Infection must be sought with vigor and treated as in other patients.

Peculiar lesions resembling erythema nodosum may develop on the shins; they are hot and painful but are not, in my experience, due to infection. They resolve spontaneously perhaps more quickly if the patient stays off his feet and applies hot compresses. The underlying periosteum is not elevated; I think they may be an adult equivalent of sickle cell dactylitis.

A variety of orthopedic procedures are available for treatment of aseptic necrosis of the femoral head (26). In recent years, we have avoided surgery until patients are unable to remain ambulatory on crutches; then we perform total hip replacement. Results range from outstanding to fair, but all patients are now able to walk either unaided or with a cane, and we have not yet experienced the disturbing complications described by others (27).

Retinopathy

Sickle retinopathy has received increasing attention in recent years, and an excellent recent review is available (28). As with many complications, eye problems become more prevalent as patients get older (29); but in contrast to the usual situation, morbidity is higher in Hb SC disease than in sickle cell anemia (30).

Feeder vessels to neovascular "sea fans" can be visualized by fluorescein angiography. They are usually occluded by xenon arc or argon laser photocoagulation, but diathermy or cryopexy is also used. Treatment can be followed by vitreous hemorrhage, ischemia of the peripheral retina, and formation of new neovascularizations (31). If retinal detachment occurs, scleral buckling can sometimes arrest the process. Partial exchange transfusion is probably more clearly indicated in eye surgery than in most other operative procedures (32).

Renal Disease

In general. renal hematuria stops spontaneously within a few days, and any treatment instituted is given credit for success. Lavage of the renal pelvis with silver nitrate, infusion of bicarbonate or distilled water, and administration of diuretics have all been advocated, and none has been consistently effective. Inhibition of urokinase by epsilon-amino-caproic acid (EACA) may lead to formation of clots in the ureter rather than cessation of bleeding. We always force fluids (to prevent clotting in the ureter), give iron (unless the patient is iron-loaded), and try to reassure him that he will not bleed to death. Such reassurance obviously cannot be given until the physician is sure the patient does not have a pyogenic or granulomatous infection, or cancer. Since hematuria tends to recur, the workup for such conditions need not be as thorough during later episodes.

Hyperuricemia is not infrequent in sicklers (33), and may be best left untreated (34). Some patients develop secondary gouty arthritis; colchicine and indomethacin are helpful in the acute attack and probenecid can be used to increase urate excretion. I have not used allopurinol because of its potential toxicity (35), but it might be of value in some patients.

PREVENTION OF VASO-OCCLUSIVE CRISES

Chronic Transfusion Therapy

In severly incapacitated patients one may wish to continue transfusions for months or years—if the patient's veins are adequate. Our attempts to create "new" veins by construction of internal arteriovenous fistulas have been miserable failures—probably because we waited until no veins were left for the surgeon to use. On a chronic transfusion program, patients must be made aware of their own responsibility to make and keep appointments for out-patient transfusion. They must also be made aware of the risks of chronic transfusion, which include hepatitis, mismatched blood, allosensitization, and iron overload. If the patient is an "antibody-former," and makes several antibodies, later transfusions may be impossible, even when urgently needed (36). Allosensitization often becomes manifest as a delayed transfusion reaction. When one occurs, measurements of the % hemoglobin S in the blood provides evidence that only transfused cells were hemolysed. Unless patients have symptoms of air hunger or angina, no attempt should be made to transfuse them, since their marrows usually respond to the stress promptly.

Iron overload is much less of a problem in sicklers than in patients with thalassemia. I have encountered two patients with questionable transfusional hemochromatosis (our autopsy files contain a few more), and clear-cut findings have been described in at least one other patient (37). Removal of one unit of blood at each transfusion slows the rate of iron accumulation, and subcutaneous infusions of desferioxamine may be useful in well-motivated patients (38). If patients on chronic transfusion programs are receiving intravenous fluids for any reason, it would seem wise to include large amounts of the drug in the IV fluid (39). Ascorbic acid supplementation increases iron excretion, but carries risks that may rule against its use (40). Strong tea (without milk) (41) may serve to decrease absorption of food iron in transfused patients.

Lowering the MCHC

Since sickling depends upon the intracellular hemoglobin concentration (41), Lincoln et al (42) suggested that a lowered MCHC might be beneficial, and Guy et al (43, 44) tried it by infusing hypotonic NaCl. More recently, Rosa and his coworkers (45) induced red cell swelling by lowering serum sodium to 120 meq/liter. Preliminary results were encouraging, but it is not clear that such treatment would be effective in the clinic, and there has been no controlled trial, at least so far.

Chemotherapy

In a recent comprehensive work, Dean & Schechter (46) reviewed potential chemotherapeutic agents (46). A few new agents have been proposed since then (47–49), but there have been no startling discoveries.

In general, proposed agents can be divided into covalent and nocovalent modifiers of hemoglobin S. Covalent modifiers are those that bind (more or less) irreversibly to hemoglobin; non-convalent agents are bound reversibly. Sodium cyanate (50) is a covalent modifier; urea (51) and carbon monoxide (52) are noncovalent agents. Agents of the former type seem a more practical form of treatment since they could be administered intermittently.

My own experience with cyanate makes me dubious that any agent which must be given systemically will be so specific in its targeting that it will modify only hemoglobin and no other important protein. Even a little change in a tissue that turns over slowly, like lens or nerve, could eventually have devastating effects. On the other hand, extracorporeal treatment of red cells with a covalently bound agent might really work. Tests of an extracorporeal carbamylator are (at last) in progress, and the prototype could be fairly easily modified for use with other agents (53). It is not clear whether such machines will ever be generally available because of their cost.

For the Future

Bone marrow transplantation, of necessity not from an identical twin, would probably do more harm than good. Injection of β genes into red cell precursors might have almost the same risk, since a significant portion of the patient's unaltered erythroid precursors would have to be ablated. A more acceptable approach would involve administering a drug or hormone that "switched" hemoglobin symthesis back to production of hemoglobin F (54). No effective agent has been proposed, and some but not all (55) recent evidence suggests that substantial amounts of fetal hemoglobin would be needed.

CONCLUSION

Sickle cell disease (as contrasted to sickle cell crises) could be prevented if conception and birth of new patients were prevented. Effective and cheap ways are available to diagnose the carrier state, hordes of counselors have been trained by federal programs, and elegant prenatal diagnosis is becoming more available. But it seems unlikely that such programs will have more than a minor effect in preventing sickle cell disease. If prospective parents seek genetic counseling, the physician must anticipate their "need to know," and their ability to assimilate information. Data should be given slowly and carefully, for emotionally involved couples may not hear what we are saying. It is incumbent upon us to go over or around the barriers and teach them enough so they can make their own decisions.

Literature Cited

1. Goodman, L. S., Gilman, A. 1970. *The pharmacological Basis of Therapeutics,* pp. 237–313. New York: Macmillan. 4th Ed.
2. Mansky, P. A. 1978. Opiates: human pharmacology. In *Drugs of Abuse, Handbook of Psychopharmacology,* ed. L. L. Iverson, S. D. Iverson, S. H. Snyder, 12:95–140. New York: Plenum
3. Laszlo, J., Obenour, W., Saltzman, H. A. 1969. Effects of hyperbaric oxygenation on sickle syndromes. *South. Med. J.* 62:453–56
4. Jasinski, D. R. 1977. Assessment of the abuse potentiality of morphine-like drugs (methods used in man). In *Drug Addiction. I. Morphine, Sedative Hypnotic and Alcohol Dependence,* ed. W. R. Martin, pp. 239–58. Berlin: Springer
5. Black, R. G. 1975. The chronic pain syndrome. *Surg. Clin. North Am.* 55: 999–1011
6. Keats, A. S., Telford J., Korosu, Y. 1961. "Potentiation" of meperidine by promethazine. *Anesthesiology* 22:34–41
7. Charache, S., Scott, J. C., Charache, P. 1979. "Acute chest syndrome" in adults with sickle cell anemia: microbiology, treatment and prevention. *Arch. Intern. Med.* 139:67–69
8. Ammann, A. J., Addiego, J., Wara, D. W., Lubin, B., Smith, W. B., Mentzer, W. C. 1977. Polyvalent pneumococcal-polysaccharide immunization of patients with sickle cell anemia and patients with splenectomy. *N. Engl. J. Med.* 297:897–900
9. Overturf, G. D., Field, R., Edmonds, R. 1979. Death from type 6 pneumococcal septicemia in a vaccinated child with sickle cell disease. *N. Engl. J. Med.* 300:143
10. Borgono, J. M., McLean, A. A., Vella, P. P., Woodhour, A. F., Canepa, I., Davidson, W. L. Hilleman, M. R. 1978. Vaccination and revaccination with polyvalent pneumococcal polysaccharide vaccines in adults and infants. *Proc. Soc. Exp. Biol. Med.* 157:148–54
11. Ahronheim, G. A., Reich, B., Marks, M. I. 1979. Penicillin-insensitive pneumococci. *Am. J. Dis. Child.* 133:187–91
12. Stockman, J. A., Nigro, M. A., Mishkin, M. M., Oski, F. A. 1972. Occlusion of large cerebral vessels in sickle-cell anemia. *N. Engl. J. Med.* 287:846–49
13. Powars, D., Wilson, B., Imbus, C., Pegelow, C., Allen, J. 1978. The natural history of stroke in sickle cell disease. *Am. J. Med.* 65:460–71
14. Cheatham, M. L., Brackett, C. E. 1965. Problems in management of subarachnoid hemorrhage in sickle cell anemia. *J. Neurosurg.* 23:488–93
15. Russell, M. O., Goldberg, H. I., Reis, L., Friedman, S., Slater, R., Reivich, M., Schwartz E. 1976. Transfusion therapy for cerebrovascular abnormalities in sickle cell disease. *J. Pediatr.* 88:382–87
16. Close, R. A., Buchheit, W. A. 1977. The management of ruptured intracranial aneurysm in sickle cell anemia. *J. Neurosurg.* 47:761–65
17. Green, T. W., Conley, C. L., Berthrong, M. 1953. The liver in sickle cell anemia. *Bull. Johns Hopkins Hosp.* 92:99–127
18. Rosenblate, H. J., Eisenstein, R., Holmes, A. W. 1970. The liver in sickle cell anemia. *Arch. Pathol.* 90:235–45
19. Cameron, J. L., Maddrey, W. C., Zuidema, G. D. 1971. Biliary tract disease in sickle cell anemia. *Ann. Surg.* 174:702–10
20. Morrison, J. C., Wiser, W. L. 1976. The effect of maternal partial exchange transfusion on the infants of patients with sickle cell anemia. *J. Pediatr.* 89: 286–89
21. Morrison, J. C., Wiser, W. L. 1976. The use of prophylactic partial exchange transfusion in pregnancies associated with sickle cell hemoglobinopathies. *Obstet. Gynecol.* 48:516–20
22. Cunningham, F. G., Pritchard, J. A. 1979. Prophylactic transfusions of normal red blood cells during pregnancies complicated by sickle cell hemoglobinopathies. *Am. J. Obstet. Gynecol.* 135:994–1003
23. Charache, S., Scott, J., Niebyl, J., Bonds, D. 1980. Management of sickle cell disease in pregnant patients. *Obstet. Gynecol.* 55:407–10

24. Schumacher, H. R., Andrews, R., McLaughlin, G. 1973. Arthropathy in sickle-cell disease. *Ann. Intern. Med.* 78:203–11
25. Lutzker, L. G., Alavi, A. 1976. Bone and marrow imaging in sickle cell disease: diagnosis of infarction. *Semin. Nucl. Med.* 6:83–93
26. Chung, S. M. K., Alavi, A., Russel, M. O. 1978. Management of osteonecrosis in sickle-cell anemia and its genetic variants. *Clin. Orthoped.* 130: 158–74
27. Gunderson, C., D'Ambrosia, R. D., Shoji, H. 1977. Total hip replacement in patients with sickle-cell disease. *J. Bone Joint Surg.* 59A:760–62
28. Goldberg, M. F. 1976. Retinal vaso-occlusion in sickling hemoglobinopathies. *Birth Defects: Orig. Art. Ser.* 12:475–515
29. Condon, P. I., Sergeant, G. R. 1976. Ocular findings in elderly cases of homozygous sickle-cell disease in Jamaica. *Br. J. Ophthalmol.* 60:361–64
30. Hannon, J. 1956. Vitreous hemorrhage associated with sickle cell–hemoglobin C disease. *Am. J. Ophthalmol.* 42: 707–12
31. Goldbaum, M. H., Galinos, S. O., Apple, D., Asdourian, G. K., Nagpal, K., Jampol, L., Woolf, M. B., Busse, B. 1976. Acute choroidal ischemia as a complication of photocoagulation. *Arch. Ophthalmol.* 94:1025–35
32. Goldberg, M. F. 1979. Sickled erythrocytes, hyphema, and secondary glaucoma. *Ophthalmic Surg.* 10:17–123
33. Diamond, H. S., Meisel, A. D., Holden, D. 1979. The natural history of urate overproduction in sickle cell anemia. *Ann. Intern. Med.* 90:752–57
34. Liang, M. H., Fries, J. F. 1978. Asymptomatic hyperuricemia: the case for conservative management. *Ann. Intern. Med.* 88:666–70
35. Young, J. L., Boswell, R. B., Nies, A. S. 1974. Severe allopurinol hypersensitivity. Association with thiazides and prior renal compromise. *Arch. Intern. Med.* 134:553–58
36. Orlina, A. R., Unger, P. J., Koshy, M. 1978. Post-transfusion alloimmunization in patients with sickle cell disease. *Am. J. Hematol.* 5:101–6
37. Buja, L. M., Roberts, W. C. 1971. Iron in the heart. Etiology and clinical significance. *Am. J. Med.* 51:209–21
38. Propper, R. D., Shurin, S. B., Nathan, D. J. 1976. Reassessment of the use of desferrioxamine B in iron overload. *N. Engl. J. Med.* 294:1421–23
39. Blume, K. G., Beutler, E., Chillar, R. K., Fahey, J. L., Sharkoff, D., Zia, P. K. 1978. Continuous intravenous deferoxamine infusion: treatment of secondary hemochromatosis in adults. *J. Am. Med. Assoc.* 239:2149–51
40. Nienhuis, A. 1977. Safety of intensive chelation therapy. *N. Engl. J. Med.* 296: 114
41. de Alarcon, P. A., Donovan, M.-E., Forbes, G. B., Landaw, S. A., Stockman, J. A. III. 1979. Iron absorption in the thalassemia syndromes and its inhibition by tea. *N.Engl. J. Med.* 300:5–8
42. Greenberg, M. S., Kass, E. H., Castle, W. B. 1957. Studies on the destruction of red blood cells. XII. Factors influencing the role of S hemoglobin in the pathologic physiology of sickle cell anemia and related disorders. *J. Clin. Invest.* 36:833–43
43. Lincoln, T. L., Aroestry, J., Morrison, P. 1973. Iron deficiency anemia and sickle cell disease: a hypothesis. *Lancet* 2:260–73
44. Guy, R. B., Gavrilis, P. K., Rothenberg, S. P. 1973. In vitro and in vivo effect of hypotonic saline on the sickling phenomenon. *Am. J. Med. Sci.* 266: 267–77
45. Rosa, R. M., Bierer, B., Thomas, R. Stoff, J., Kruskall, M., Robinson, S., Bunn, H. F., Epstein, F. H. 1980. Prevention of sickle cell crises by sustained hyponatremia. *Clin. Res.* 28:549A
46. Dean, J., Schechter, A. N. 1978. Sickle cell anemia: molecular and cellular bases of therapeutic approaches. *N. Engl. J. Med.* 299:752–63, 804–11, 863–70
47. Kark, J. A., Kale, M. P., Tarassoff, P. G., Woods, M., Lessin, L. S. 1978. Inhibition of erythrocyte sickling in vitro by pyridoxal. *J. Clin. Invest.* 62:888–91
48. Nigen, A. M., Manning, J. M. 1978. Effects of glyceraldehyde on the structural and functional properties of sickle erythrocytes. *J. Clin. Invest.* 61:11–19
49. Walder, J. A., Zaugg, R. H., Walder, R. Y., Steele, J. M., Klotz, I. M. 1979. Diaspirins that cross-link β chains of hemoglobin: *bis*(3,5-dibromosalicyl) succinate and *bis*(3,5-dibromosalicyl) fumarate. *Biochemistry* 18:4265–70
50. Harkness, D. R., Roth, S. 1975. Clinical evaluation of cyanate in sickle cell anemia. *Prog. Hematol.* 9:157–84
51. Cooperative Urea Trials Group. 1974. Clinical trials of therapy for sickle cell vaso-occlusive crises: treatment of

sickle cell crises with urea in invert sugar—a controlled trial. *J. Am. Med. Assoc.* 228:1120–24

52. Beutler, E. 1975. The effect of carbon monoxide on red cell life span in sickle cell disease. *Blood* 46:253–59
53. Babb, A. L., Scribner, B. H., Uvelli, D. A., Fry, D. A., Fry, D. L., Agodoa, L. C. 1979. An extracorporeal system that modifies hemoglobin in patients with sickle cell anemia. *Artif. Org.* 3:62–72
54. Pembrey, M. E., Wood, W. G., Weatherall, D. J., Perrine, R. P. 1978. Fetal haemoglobin production and the sickle gene in the oases of eastern Saudi Arabia. *Br. J. Haematol.* 40:415–29
55. Powars, D. R., Schroeder, W. A., Weiss, J. N., Chan, L. S., Azen, S. P. 1980. Lack of influence of fetal hemoglobin levels or erythrocyte indices on the severity of sickle cell anemia. *J. Clin. Invest.* 65:732–40

Ann. Rev. Med. 1981. 32:207–212
Copyright © 1981 by Annual Reviews Inc. All rights reserved

OUTLOOK FOR ACUTE LYMPHOCYTIC LEUKEMIA IN CHILDREN IN 1982[1]

❖7392

Joseph V. Simone, M.D.

St. Jude Children's Research Hospital, Memphis, Tennessee 38101

INTRODUCTION

The outlook for children with acute lymphocytic leukemia (ALL) must be viewed from the perspective of history and a critical assessment of current therapeutic efficacy. Modern treatment of ALL began with the discovery and successful use of aminopterin by Dr. Sidney Farber and his colleagues after the second world war. This was followed over the next 15 years by the introduction of a series of effective agents—corticosteroids, mercaptopurine, and vincristine. Discovery and application of new and effective agents has continued, but at a far less spectacular pace. Of the established major agents, only asparaginase and the anthracyclines were discovered after 1960. The effective use of combinations of agents began in the late 1950s with the establishment of standards of therapeutic response and of specific experimental designs to test a variety of simultaneous and sequential drug combinations. This era also has seen the introduction of two critical new approaches—the determined aim for cure of ALL instead of palliation and the conceptual development and successful application of therapy for prevention of CNS leukemia. The decade of the 1970s has largely been a time of therapeutic refinement—spreading the concepts and skills of modern therapy, trying permutations of drug schedule and central nervous system (CNS) therapy, and attempting to improve results and reduce late effects by using the clinical prognostic category to determine the intensity of

---

[1] Supported by Grants CA08480, CA20180, CA21765, CA23944 from the National Cancer Institute and by ALSAC.

0066-4219/81/0401-0207$01.00

therapy. Two promising innovations in this era have been the recognition of leukemia cell features (mainly immunological) and their relationship to prognosis, and the clinical trials of bone marrow transplantation for treatment of leukemia with curative intent.

The treatment of ALL continues to evolve since optimal therapy has not yet been devised. Current therapy is toxic, prolonged, and, most important, a failure in over one half of all patients. An even higher proportion of adults and other prognostic subsets of patients fail with current therapy. In a real sense, therefore, "standard" therapy for ALL does not exist; all current therapy regimens are (or should be) experimental. Descriptions of current therapy should be read with that caveat in mind.

Current Therapy

REMISSION INDUCTION THERAPY Therapy aimed at inducing remission is effective. The addition of asparaginase or an anthracycline to the basic combination of vincristine and prednisone induces remission in all but 5% of patients with ALL. Since there was evidence that, to a point, the intensity of remission induction therapy influenced the duration of remission (1), a current study at St. Jude Hospital was designed to test whether administration of four agents during remission induction influenced the duration of remission; so far, patients given the fourth drug have no longer remissions. It would appear, therefore, that a three drug regimen is optimal with presently available agents.

PREVENTIVE CNS THERAPY Attention of investigators is now most closely focused on preventive CNS therapy. Although administration of 2400 rads of cranial radiation with intrathecal methotrexate remains the standard by which all other forms of CNS therapy must be measured, concern for potential late effects of this therapy (2, 3) has led to the development of alternate forms of CNS therapy. The longest successful experience is with intrathecal methotrexate and aggressive systemic therapy (4), which appears to be as effective as irradiation. Other studies of intrathecal therapy, reduced dosage irradiation, and intermediate dosage methotrexate infusions by vein (5–7) are well underway. It is likely that a variety of forms of CNS prophylaxis will be effective, but it will take time to determine the relative toxicity of each and whether any has specificity for a subgroup of patients.

CONTINUATION THERAPY Continuation therapy during remission has undergone numerous modifications. However, the basic treatment standard continues to be a combination of mercaptopurine and methotrexate. Although promising additions are under study, there is no convincing evidence that the addition of other agents during remission significantly prolongs its duration (1, 7–9), provided both agents are given in maximum-tolerated dosage. This phase of therapy has received the least attention in recent years and yet is responsible for the vast majority of treatment failures in ALL. Most failures are now due to bone marrow relapse that is attributable to the development of resistance to the agents given during remission. Above all, this is the phase of therapy that most needs a fresh approach and the courageous trial of new ideas. At St. Jude we are reexamining the use of elective sequential therapy to prevent emergence of resistant leukemia cells. Instead of single agents, we are testing the value of pairs of agents each given for several months. This study has just started and we have no data to report.

Exciting results are being obtained by Reihm and co-workers (10) in a series of West German studies. They give very intensive multiagent therapy to all patients during the first 8–12 weeks of therapy. Therapy is so aggressive during that time that skillful and intensive supportive care is required. Subsequent modifications of these studies, not yet in print, give preliminary indications that patients with "poor risk" features may do as well as those with "standard risk" features at diagnosis. If this is borne out with further follow-up, the advance would be most significant and justify emulation and confirmation elsewhere.

CESSATION OF THERAPY After 28–36 months of initial complete remission, therapy is ceased at St. Jude; subsequently, 75–80% of patients remain in remission indefinitely. As a result of this observation, we have developed an operational definition of cure for this disease, that is, that the patient remain in remission for four years after cessation of therapy (11). Longer durations of therapy are being tested by the cooperative groups. Reihm is currently exploring a 24-month therapy period. Although the 20–25% relapse rate in our studies is not acceptable, we believe prolonging the same therapy most likely will merely shift the time of relapse of these patients without reducing the number of relapses. Because of the relatively high frequency of testicular relapse in boys after cessation of therapy, we now routinely perform bilateral testicular biopsies at the time therapy is to be discontinued.

BONE MARROW TRANSPLANTATION The results of bone marrow transplantation for ALL have generated new interest with the promising results of transplanting patients during remission (12). By transplanting patients during a second or subsequent remission, many of the early deaths due to infectious and hemorrhagic complications may be avoided. If the current trend continues, the proportion of long-term survivors will be as high as 30–50% compared to the 7–15% among patients transplanted mainly during relapse.

If technical and scientific advances continue toward development of more specific and sensitive antibodies, successful autologous transplantation is likely to become a potent therapeutic tool. One would achieve remission with conventional therapy, withdraw a quantity of the patient's marrow known to harbor a small percentage of leukemia cells, and remove those cells with a monospecific antibody. After conventional pretransplant therapy aimed at eradication of residual leukemia, the patient would be given his own "clean" marrow cells. This approach has two major advantages. First, there is no need for matching tissue types. Currently, only about one third of patients have a suitable donor. Second, there is no graft-versus-host reaction which can be chronic and debilitating.

Leads for the Future

During the next decade we may expect further refinement of the activities noted above: more information on the effectiveness and side-effects of the various forms of preventive CNS therapy now being tested; epidemiological studies of late effects of all kinds in the increasing number of long-term survivors; completion of current studies of the optimal duration of therapy by the Children's Cancer Study Group and the Cancer and Leukemia Group B; and continued studies of ways to use current agents more effectively. The latter remains an important exercise—it is remarkable that we are still seeking the optimal schedule and dosage for the use of methotrexate in ALL more than 30 years after its introduction. There will be continued attempts to improve the results of treatment of patients with poor prognostic features. Thus far, the rationale has been that this more resistant form of the disease requires more intensive therapy with more drugs and higher dosages. However, this approach has not improved the results. More effective therapy may depend upon the emergence of useful clues from studies of the biological basis of resistance (13); this is more likely to occur now because of the relative ease of identification of some subclasses of leukemia cells. An empirical approach is unlikely to be successful without a radical departure from current regimens or the discovery of a new and highly effective drug. With all the focus on patients with features associated with a poor prognosis, insufficient attention has been paid to the fact that we do not understand the biological basis for the failure of patients without poor prognostic features. At least one third of "Standard risk" patients will develop a bone marrow relapse despite their designation as "common" ALL. It is critical that we learn whether these relapses are avoidable by the more optimal use of current agents or if resistance is innate.

There are two new agents with some promise for treatment of ALL. Studies of the combination of the epipodophyllotoxin VM-26 and cytosine arabinoside have shown considerable promise in patients refractory to other forms of treatment (14). The combination is so promising in fact, that it is being incorporated into our new front-line studies for ALL. Another agent that bears watching, both because of its early promise and because of its unique mode of action, is deoxycoformycin (15). This agent is an adenosine deaminase inhibitor that specifically attacks lymphoid cells with little apparent effect on other bone-marrow-derived cells. It is too early to determine whether this specificity will hold up in large trials.

The most urgent need of the 1980s, in my view, is a method for detecting minimal residual disease in the bone marrow of patients who are in remission of ALL. If we had the capacity for detecting one leukemia cell in a thousand or a million or more, we would be able to measure the effectiveness of therapy given during remission. Currently, of course, we must wait for relapse to occur and we have no clue as to when resistance may have developed. Detection of minimal residual disease is within reach for ALL because of the rapid development of our ability to raise quantities of monovalent antisera with hybridomas, and because of technology such as flow cytofluorimetry with cell separation capability.

Summary

The treatment of childhood ALL has been remarkably successful over the past 30 years of the chemotherapy era. A substantial proportion of patients, perhaps 40–50%, are apparently cured with currently available therapy. However, optimal therapy has not yet been devised because of substantial therapeutic failures in the form of relapse and unacceptable side-effects. Remission induction therapy is the most settled aspect of therapy. The addition of asparaginase or an anthracycline to prednisone and vincristine is highly effective. Preventive central nervous system therapy is under revision and study. It is likely that a variety of approaches will be equally effective. Continuation (maintenance) therapy is in the most need of revision because a majority of therapeutic failures are due to bone marrow relapse—the failure of systemic therapy to prevent emergence of resistant leukemia cells. The optimal duration of therapy has not been established, the relapse rate after cessation of therapy lasting 2–3 years currently being 20–25%. Bone marrow transplantation during remission shows promise as a therapeutic modality, particularly if autologous techniques are successful. The most promising development, however, is our increasing understanding of biological subpopulations of leukemia cells. This ultimately may help us develop more effective and specific therapy for childhood ALL.

Literature Cited

1. Simone, J. V. 1976. Factors that influence haematological remission duration in acute lymphocytic leukaemia. *Br. J. Haematol.* 32:465–72
2. Eiser, C., Lansdown, R. 1977. Retrospective study of intellectual development in children treated for acute lymphoblastic leukemia. *Arch. Dis. Child.* 52:525–29
3. Peylan-Ramu, N., Poplack, D. G., Pizzo, P. A., Adornato, B. T., DiChiro, G. 1978. Abnormal CT scans of the brain in asymptomatic children with acute lymphocytic leukemia after prophylactic treatment of the central nervous system with radiation and intrathecal chemotherapy. *N. Eng. J. Med.* 289:815–18
4. Haghbin, M., Tan, C. T. C., Clarkson, B. D., Mike, V., Burchenal, J. H., Murphy, M. L. 1975. Treatment of acute lymphoblastic leukemia in children with "prophylactic" intrathecal methotrexate and intensive systemic therapy. *Cancer Res.* 35:807–11
5. Nesbit, M. E., Sather, H. N., Robison, L. L., D'Angio, G. J., Littman, P., Donaldson, M., Hammond, G. D. 1979. Presymptomatic CNS treatment in childhood acute lymphoblastic leukemia: comparison between 1800 and 2400 rads. *Proc. Am. Assoc. Cancer Res.* 20:343
6. Henze, G., Langermann, H. J., Lampert, F., Neidhardt, M., Reihm, H. 1979. Die Studie zur Behandlung der akuten lymphoblastischen Leukaemie 1971–1974 der deutschen Arbeitsgemeinschaft fur Leukaemieforschung und Behandlung im Kindesalter e.V. Analyze der prognostischen Bedeutung von Initialbefunden und Therapievarianten. *Klin. Paediatr.* 191:114–26
7. Freeman, A. E., Wang, J. J., Sinks, L. F. 1977. High-dose methotrexate in acute lymphocytic leukemia. *Cancer Treat. Rep.* 61:727–31
8. Aur, R. J. A., Hustu, H. O., Verzosa, M. S., Wood, A., Simone, J. V. 1973. Comparison of two methods of preventing central nervous system leukemia. *Blood* 42:349–57
9. Aur, R. J. A., Simone, J. V., Verzosa, M. S., Wood, A., Stagner, A., Mason, C. 1978. Childhood acute lymphocytic leukemia. Study VIII. *Cancer* 42: 2123–34
10. Reihm, H., Gadner, H., Welte, K. 1977. Die West-Berliner Studie zur Behandlung der akuten lymphoblastichen Leukaemie des Kindeserfahrungsbericht nach 6 Jahren. *Klin. Paediatr.* 189:89–102
11. George, S. L., Aur, R. J. A., Mauer, A. M., Simone, J. V. 1979. A reappraisal of the results of stopping therapy in childhood leukemia. *N. Engl. J. Med.* 300:269–74
12. Thomas, E. D., Sanders, J. E., Flournay, N., Johnson, F. L., Buckner, C. D., Clift, R. A., Fefer, A., Goodell, B. W., Storb, R., Weidern, P. L. 1979. Marrow transplantation for patients with acute lymphoblastic leukemia in remission. *Blood* 54:468–76
13. Ohnuma, T., Arkin, H., Minowada, J., Holland, J. F. 1978. Differential chemotherapeutic susceptibility of human T-lymphocytes and B-lymphocytes in culture. *J. Natl. Cancer Inst.* 60:749–52
14. Rivera, G., Aur, R. J. A., Dahl, G. V., Avery, T., Pratt, C. B. 1979. VM-26 and cytosine arabinoside (ara-C) combination chemotherapy for initial induction failures in childhood lymphocytic leukemia. *Proc. Am. Soc. Clin. Oncol.* 20:370
15. Smyth, J. F., Chassin, M. M., Harrap, K. R., Adamson, R. H., Johns, D. G. 1979. 2-Deoxycoformycin (DCF): Phase I trial and clinical pharmacology. *Proc. Am. Assoc. Cancer Res.* 20:47

Ann. Rev. Med. 1981. 32:213–20
Copyright © 1981 by Annual Reviews Inc. All rights reserved

CARDIAC TRANSPLANTATION ❖7393

Sharon A. Hunt, M.D. and Edward B. Stinson, M.D.

Department of Medicine (Cardiology Division) and Department
of Cardiovascular Surgery, Stanford University School of Medicine,
Stanford, California 94305

Introduction

The first human cardiac transplant operation at Stanford in January 1968 was preceded by over a decade of laboratory work refining the surgical technique of orthotopic transplantation and the details of immunosuppressive treatment necessary for allograft survival (2, 3, 6). In the 12 years since then, over 400 transplant procedures have been performed worldwide. The largest series from a single institution comes from the Stanford program. Since current survival rates for cardiac allograft recipients reported from Stanford (1) are equal to those reported for organ survival in cadaver renal transplant recipients (4), cardiac transplantation is becoming increasingly accepted as a therapeutic procedure, and other transplant programs can be expected to develop and flourish in years to come. We describe here several aspects of the Stanford cardiac transplant program, including criteria for recipient and donor selection, anatomy relevant to the surgical technique, current methods of immunosuppression, diagnosis and treatment of rejection, and the variety and incidence of other complications. Cardiac retransplantation has become a standard treatment for several complications and we describe our experience with that procedure as well.

Recipient Selection

Patients with end-stage heart disease irremediable by any more conventional medical or surgical treatment are considered potential candidates for transplantation if several other criteria are met. These criteria were developed empirically to select a subset of patients most likely to benefit from the procedure. They include age under 55 years, freedom from any active infection or unresolved pulmonary infarct, freedom from any other systemic disease that would separately limit survival, and the lack of inordi-

0066-4219/81/0401-0213$01.00

nately elevated pulmonary vascular resistance. The last criterion is included because our early experience showed that a normal donor right ventricle cannot increase its external work load acutely enough to cope with high resistance to outflow. Psychosocial criteria are less easy to quantitate, but are very important determinants of postoperative quality of life as well as survival. A history of good compliance with medical regimens, general psychosocial stability, and a positive surrounding emotional support structure are all factors that are weighed in evaluating potential transplant recipients.

Virtually all patients have presented with functional class IV NYHA cardiac disability. The majority have manifested primarily signs and symptoms of congestive heart failure and low cardiac output; several, however, have also exhibited intractable, life-threatening arrhythmias.

The underlying heart disease in our patient group is most frequently coronary artery disease (49% of patients). Many patients (43%) have cardiomyopathy of other etiology (idiopathic, viral, postpartum, familial), several (3%) have had combined cardiomyopathy and valvular heart disease, and one patient had a posttraumatic left ventricular aneurysm. Two patients to date have presented with inoperable congenital heart disease.

Previous heart surgery is not a contraindication to selection since patients having had such procedures, either valve replacement or coronary artery bypass grafting, do not present major technical problems at the time of transplantation; furthermore, our experience has shown that a previous cardiac operation that failed to provide long-term therapeutic benefit does not jeopardize, and may improve, expectations for survival after transplantation.

On the basis of the above criteria, 250 patients have been accepted as cardiac transplant recipients as of 1 May 1980. Of these, 61 died before an appropriate donor became available and 188 of the remainder have undergone transplantation (one is currently waiting). The 250 selected recipients represent 10–20% of the total number of patients referred to Stanford for consideration of transplantation.

Donor Selection

Patients who meet the medical and legal criteria of brain death, are less than 35 years old, have no history of known cardiac disease, and have a normal cardiac examination and electrocardiogram, are considered appropriate cardiac donors. The majority of donors in the Stanford program died as a result of closed head injury, and the remainder suffered either gunshot wounds of the head or spontaneous intracranial hemorrhage.

Several aspects of donor and recipient matching preoperatively are important. Comparability in body weight, ABO blood group compatibility, and the absence of a positive lymphocyte crossmatch between donor and recipient are all required. Histocompatibility (HLA) matching is not used prospectively since our experience has not shown any correlation between early postoperative results and number of antigenic mismatches. One exception is that mismatches of the HLA-A2 antigen now appear to be associated with an increased incidence of late postoperative graft atherosclerosis. Matching for this antigen has been required only since January 1979 and the effect of this change in selection criteria cannot be assessed for another one or two years.

In the initial years of the program, donor hearts were removed at Stanford. Since 1977, a program of distant heart procurement has been instituted, with donor hearts removed at other institutions after hyperkalemic arrest, preserved by immersion in 4°C saline, and transported to the waiting recipient at Stanford. Ischemic times have not exceeded 3½ hours and all such grafts have functioned satisfactorily. The success of this method provides a new and slightly less expensive source of donor organs, the availability of which has heretofore been a major impediment to expansion of the program.

Operative Technique

Techniques for orthotopic heart transplantation have changed little since the original description of successful heart transplantation in the laboratory in 1960. After institution of cardiopulmonary bypass, the recipient heart is excised by division of the atria at their midlevel and of the great vessels immediately above the semilunar valves. This maneuver leaves *in situ* the posterior walls of the recipient right and left atria to which are anastomosed the corresponding structures of the donor heart; the sinoatrial node in the residual recipient right atrium is retained. Care is taken to preserve the integrity of the donor sinoatrial node, and although rhythms of apparently low atrial origin are common in the early postoperative period, nearly all recipients eventually exhibit regular sinus rhythm.

Immunosuppression

With currently used techniques, recipients of cardiac allografts require indefinite immunosuppression to control host immune responses that would otherwise result in rejection of solid organ allografts. The immunosuppressive regimen is most intense in the early postoperative period when rejection episodes are most frequent; immunosuppression tapers to lower levels over the subsequent years. Increases in the chronic or baseline level of immunosuppression, necessary for treatment of acute rejection episodes, are described below. Early intensive treatment consists of a "triple" regimen of high-dose corticosteriods, azathioprine, and anti-human thymocyte globulin (ATG) produced in rabbits immunized against human thymocytes. The ATG is discontinued after the first 3–4 weeks and corticosteroid doses are gradually tapered from an initial 1.5 mg/kg/day of prednisone down to 1.0 mg/kg/day in 2–3 weeks and to a minimum of 0.25 mg/kg/day in the first 1–2 years. Azathioprine administration is continued indefinitely at the highest dose compatible with adequate peripheral white blood cell and platelet counts; the maximum dose is approximately 200 mg/day.

Survival Rates

Patient survival, calculated by the actuarial method, for the entire Stanford series of patients is shown in Figure 1. Subgrouping the patients into cohorts operated on before and since January 1, 1974, illustrates a marked improvement in survival expectancy for more recent patients, with survival rates of 62, 58, and 54% at one, two, and three years, respectively, as contrasted to corresponding rates of 42, 35, and 28% for the earlier group. This improvement can be attributed to initiation of use of the current type of ATG in 1974 and development of the transvenous endomyocardial biopsy technique to diagnose and follow acute graft rejection, as well as advances in the early recognition and available treatment of infectious complications. These recent survival rates are entirely comparable to those reported for organ survival for most series of cadaver renal transplants (4).

The only subgroup exhibiting clearly superior survival rates is that of younger patients. The one-year survival rate in patients under 40 years old at the time of operation is over 70%. The type of original cardiac disease does not influence postoperative survival.

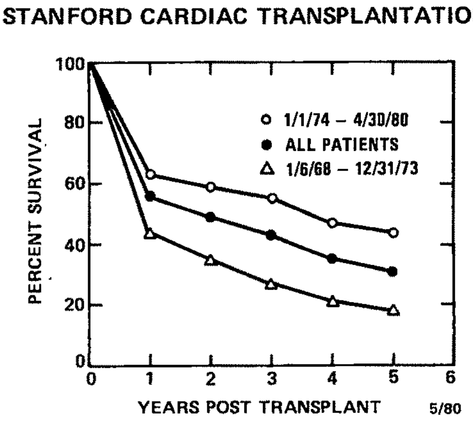


Figure 1 Actuarial survival curves for patients in the Stanford cardiac transplantation program.

Rejection

INCIDENCE Acute rejection of the cardiac allograft occurs at some time postoperatively in approximately 90% of patients. Its frequency is greatest in the early postoperative period, occurring on the average once per 35 patient-days, and declines over time, occurring once per 748 patient-days after the first year. The severity of acute rejection is quite variable histologically, but unless and until it is quite advanced (and frequently irreversible), it does not result in signs or symptoms of congestive heart failure or low cardiac output. Severe acute rejection led to eight deaths in the early years of the Stanford program, but such deaths are fortunately rare in recent experience.

DIAGNOSIS Clinical signs of acute allograft rejection consist of the appearance of an early diastolic sound as a manifestation of decreased left ventricular compliance and the development of an irregular pulse due to a variety of atrial arrhythmias (atrial premature contractions, atrial flutter, atrial fibrillation). Ventricular arrhythmias do not characteristically occur with acute rejection. A generalized decrease in EKG voltage is the most sensitive and earliest noninvasive indicator of acute rejection, but false positive shifts are not infrequent. A method has been devised and validated in the laboratory for summing the amplitudes of the total QRS deflections in leads I, II, III, V_1, and V_6 using a carefully calibrated EKG machine with constant precordial lead placement assured by placing tattoos on the V_1 and V_6 positions. This voltage sum falls in a reasonably narrow range in a given individual; decreases of 10% or more are suggestive of acute rejection.

Histological documentation of the presence and severity of acute rejection is obtained by means of percutaneous transvenous endomyocardial biopsy in all patients managed at Stanford, thus eliminating potentially hazardous increases in immunosuppression in patients with false positive EKG or clinical signs of rejection. This procedure provides a safe means to assess objectively not only the initial process, but also to confirm its complete resolution after treatment.

Pathological changes seen with acute rejection are characteristic of classical first-set rejection and consist of cellular infiltration with predominantly mononuclear cells, interstitial and intracellular edema, and vascular congestion. Later, myocytolysis may develop.

TREATMENT Increases in immunosuppression necessary to treat acute allograft rejection are somewhat different in early and late postoperative periods. In the early period, when corticosteroid doses are already high, acute rejection is usually treated with large doses of methylprednisolone and reinstitution of or increase in ATG administration. In the later period, treatment is more individualized according to the clinical and histological severity of rejection. Treatment for the mildest episodes consists of increase of oral prednisone; the severest episodes mandate the same treatment as described for episodes early postoperatively. Severe rejection unresponsive to treatment has been treated by retransplantation in six patients to date, from 3½ to 35 months after the first transplant. Survival after a second transplant for acute rejection ranges from 15 days to 5½ years, with two patients still alive at 18 months and 5½ years after the second procedure.

Complications

INFECTIONS Infectious complications constitute a major source of morbidity and mortality in all series of organ transplantation. The Stanford series has had a total of 571 infectious episodes to date and 55% of deaths have been due to infection. A wide variety of organisms has been encountered; a large proportion are opportunistic pathogens such as *Aspergillus, Nocardia,* and *Pneumocystis.* A program of trimethoprim/sulfa prophylaxis has recently been shown to reduce the incidence of infections due to the latter two organisms (unpublished data). The majority of infections are pulmonary in origin, as shown in Table 1.

STEROID-RELATED PROBLEMS All patients receiving corticosteroids on a chronic basis are subject to a variety of related complications, including steroid-related diabetes mellitus, posterior subcapsular cataracts, and osteoporosis. In heart transplant recipients, osteoporosis is the most disabling complication. Severe back pain due to vertebral collapse led to serious disability in ten patients, and three patients required bilateral total hip replacement because of aseptic necrosis of the femoral heads.

Table 1 Stanford cardiac transplantation infectious episodes

Infection	Number of episodes	Number of patients
Pulmonary	253	123
Empyema	10	10
Septicemia	61	47
Urinary tract	36	26
Disseminated fungal	10	10
Disseminated viral	13	13
Central nervous system	20	20
Hepatitis	8	7
Retinitis	4	4
Osteomyelitis	1	1
Miscellaneous	145	95

MALIGNANCY As has been noted in other groups of chronically immunosuppressed patients, heart transplant recipients sustain an increased incidence of lymphoproliferative malignancies. To date, there have been 11 such malignancies in the Stanford series of 188 patients. Two of these consisted of acute leukemia that led rapidly to death. The solid tumors have all had a similar histologic pattern, described as lymphoblastic sarcoma, which has been classified with other immunosupression-related lymphomas (5). They tend to occur as isolated lesions in unusual extranodal locations. Two were found "incidentally" at autopsy in patients dying of other causes and one patient died of an unrelated cause before treatment could be begun. Only two patients had multicentric disease and in both the lymphoma arose in the anterior thigh muscles at the sites of previous ATG injections and spread rapidly despite radiation therapy. Both patients had undergone a second transplant. Four patients have been successfully treated with radiotherapy (two for isolated CNS lesions and two for isolated pulmonary nodules), and three are alive with no evidence of disease between six months and three years after treatment. The fourth died of an unrelated cause six months after treatment and showed no evidence of lymphoma at autopsy. Thus, lymphoma, while a serious complication, may be very treatable when it does not arise at the sites of ATG injection.

GRAFT ATHEROSCLEROSIS The complications described above are common to all recipients of organ allografts. One complication apparently unique to recipients of cardiac allografts is a seemingly accelerated form of coronary atherosclerosis. This occurs equally commonly in patients whose original heart disease was atherosclerotic or idiopathic. The exact etiology is unclear but is probably related predominantly to immunologically mediated injury of the donor coronary arterial endothelium during rejection episodes, with some modulation by "risk factors" such as elevated blood lipids, arterial hypertension, etc. This form of atherosclerosis leads to the same sequelae as the ordinary form of atherosclerotic coronary disease with one exception. The patients can develop congestive heart failure, ventricular arrhythmias, and sudden death, but they do not develop angina pectoris because of the denervated status of the allograft.

We have calculated by actuarial means that 38% of all our patients develop evidence of some degree of graft atherosclerosis by five years post-transplant. To date, nine patients in the Stanford series have developed graft atherosclerosis severe enough to be life-threatening and to require elective retransplantation. Three of these patients are still alive between 9 and 20 months after graft replacement; another lived for 16 months before succumbing to a presumed arrhythmia. Thirteen patients early in the series died of this secondary disease before the concept of retransplantation became accepted. The clinically silent appearance of this life-threatening complication provides the rationale for performing coronary arteriograms annually on heart transplant recipients and considering intervention with retransplantation before serious graft dysfunction develops.

Postoperative Rehabilitation

The cardiac allograft exhibits normal hemodynamic function under conditions of rest. In the absence of autonomic reflexes, the denervated heart relies on the Frank-Starling mechanism and on circulating catecholamines to vary its output and these mechanisms are quite sufficient to support normal physical activities.

Fully 90% of the transplant recipients, having been in NYHA functional class IV preoperatively, are able to pursue a normal, physically unrestricted life and return to the occupation or activity of their choice. None has been limited by cardiac disability.

The intensity of medical surveillance for rejection or infectious complications diminishes with time, but even the longest-term patients are seen every six weeks and return to Stanford annually for invasive evaluation.

Summary

Based on the Stanford series of 188 patients, we emphasize that orthotopic cardiac transplantation represents a reasonable therapeutic option for the patient with inoperable end-stage heart disease. Current expectations for survival equal those for organ survival in cadaver renal transplant series and offer a rehabilitation potential of over 90%. Most of the complications limiting survival are due to the currently available means of immunosuppression and, therefore, advances in the field of immunosuppression can be expected to have a major positive influence on survival rates.

Literature Cited

1. Baumgartner, W. A., Reitz, B. A., Oyer, P. E., Stinson, E. B., Shumway, N. E. 1979. Cardiac homotransplantation. *Current Problems in Surgery,* ed. M. M. Ravitch, Vol. XVI, No. 9. Chicago & London: Year Book Medical. 61 pp.
2. Lower, R. R., Dong, E., Shumway, N. E. 1965. Suppression of rejection crises in the cardiac homograft. *Ann. Thorac. Surg.* 1:645–49
3. Lower, R. R., Shumway, N. E. 1960. Studies on orthotopic transplantation of the canine heart. *Surg. Forum* 11:18
4. Monaco, A. 1979. Transplantation. In *Strauss & Welt's Diseases of the Kidney,* ed. L. E. Earley, C. W. Gottschalk, 1:463–515. Boston: Little, Brown. 3rd ed.
5. Penn, I. 1979. Tumor incidence in human allograft recipients. *Transplant. Proc.* 11(1):1047–54
6. Stinson, E. B., Griepp, R. B., Bieber, C. P., Shumway, N. E. 1972. Hemodynamic observations after orthotopic transplantation of the canine heart. *J. Thorac. Cardiovasc. Surg.* 63:344–52

Ann. Rev. Med. 1981. 32:221–31
Copyright © 1981 by Annual Reviews Inc. All rights reserved

THE PROBLEM OF OBESITY

◆7394

Faith T. Fitzgerald, M.D.

Department of Medicine, University Hospital, University of Michigan,
Ann Arbor, Michigan 48109

INTRODUCTION

Any physician who cares for patients recognizes the painful problem of obesity. It is estimated that 25–45% of the American population over 30 years of age is "overweight" (1). Doctors see obesity in the clinic, the office, the hospital, and on the street. And even if we avoid direct patient involvement by engaging in nonclinical activities, we still worry about it in ourselves and in our families.

Examination of the scientific and lay literature about obesity finds it predominantly echoing Ayer's statement of over 20 years ago: "Practically everyone will agree that obesity is the number one health problem today. . . ." (2). But are there well-established data that support this pronouncement? Is there a firm scientific foundation to the national, almost obsessive, opposition to fatness? Is the medical community, in particular, justified in its continuing efforts to promote weight reduction among the obese?

This review presents three theses for your consideration. The iconoclastic views that are suggested are hardly new, nor do they lack for substantial evidence. They are:

1. Obesity, as we commonly use the term, may be more an aesthetic and moral problem than one of physical health (3–5).
2. The therapy of obesity may be, in some circumstances, more morbid than fatness (6–11).
3. There may be some advantages, in medical and other senses, to being fat (3).

MORAL PROBLEMS

What do we mean when we say someone is obese? By some criteria, up to 60% of Americans aged 40–60 years are over their "best weight" by more

0066-4219/81/0401-0221$01.00

than 10% (12). One third of our nation's people are said to be more than 20% over their "ideal weight" (1). Yet a state cannot be called abnormal in any statistical sense if it is characteristic of so large a number of the population under study. It would seem to call for a redefinition of what is normal.

Obesity has also been defined as a state of excess body fat. Studies by cadaver analysis showed the "average" young man to possess 15% of his total body weight as fat, and the "average" young woman 25% (13). But these percentages come from a total of but eight cadavers (14), which casts doubt on their universal applicability. Moreover, this sort of analysis is obviously impractical in the doctor's office and of little aid to the individual physician trying to diagnose obesity in the individual patient. Assessment of body density, and thus of the proportion of fat to lean by the Archimedian method of immersion and water displacement (15, 16) suffers from inaccuracies of method (3, 14) and is too cumbersome and sloppy a technique to use in clinic. Isotope or chemical dilution studies of fat and volumes of distribution are best suited to research laboratories (17, 18).

Anthropomorphic measurements used in studies of large populations of fat people use skin-fold thickness, height-weight ratios, ponderal indices, and the like, all compared to arbitrary normals (14). For practical purposes, the most frequently cited criteria for normal weight refer to height-weight comparisons for each age and sex category (19). These charts are generally based on the Society of Actuaries report of 1959, which analyzed persons insured from 1935 to 1953 and showed a linear relationship between overweight and excess mortality. In this survey, the aggregate mortality for men and women more than 20% overweight was found to be about 150% that of the insured population of average weight (19).

It has been argued, however, that these insurance data are seriously flawed (20–22). Lower socioeconomic classes, which tend to be heavier, are less likely to be insured. The health awareness and performance of the newly insured population may be unusual. Height and weight estimates of this group were notoriously bad; moreover, the insurance data do not consider body weight in terms of adiposity, but only of raw weight in relation to age, height, and sex. There is no distinction between those who were heavy with muscle and those burdened with fat. These same data could be used to label body building as a major health hazard.

A recent review compares these actuarial data with six other epidemiologic studies. The results are strikingly different. There may actually be a negative relationship between body weight and mortality. The obese, in some age categories, may live *longer* than the lean (23).

Most physicians do not habitually use reference charts and probably could not quote the ideal weight for a 5 foot-5 inch woman in 2-inch heels with medium frame. Most doctors, like most lay people, diagnose obesity by eyeball analysis. Well-established experts in the field of obesity are untroubled by recommending superficial observation (12) or the "mirror test" (14)—looking at one's naked body in the mirror (24)—as a reasonable way to estimate obesity. Do the patients seem fat to us? Then fat they are. And in this method of diagnosis we are biased by culture: What an American fashion model calls chubby a Russian grandmother considers pathologically thin.

Self-perception of obesity is common in our weight-conscious society. In one sample of 500 Californians, 32% of the men and 46% of the women described themselves as too heavy (14). This is generally perceived as a perjorative, and can be potentially harmful.

Because of the cultural equation of beauty with goodness, the aesthetic aspects of overweight blend into the moral. A fat man says, "Just looking at myself in a store window makes me feel terrible. It has gotten so that I am very careful not to look by accident. It's a feeling that people have a right to hate me and hate anyone who looks as fat as me. As soon as I see myself, I feel an uncontrollable burst of hatred. I just look at myself and say, 'I hate you: you're loathesome' " (25).

It is clear from reading magazines or watching television that public derision and condemnation of fat people is one of the few remaining sanctioned social prejudices in this nation freely allowed against any group based solely on appearance. Personality profiles of obese adolescent girls show that they respond very like oppressed minorities in their acceptance of dominant values in the culture and their passive withdrawal (26). It has been documented that all of us—the general public, social workers, employers, graduate schools admissions officers, nurses, and physicians—feel negatively toward the obese (4, 27–29). Obesity is a moral crime and one of the few remaining sinful diseases: one is fat because of "weakness of the will" (30). The fat are denied jobs, promotions, educational opportunities (14), and, recently, challenged in their right to adopt a child until they lose weight (31).

MEDICAL PROBLEMS

Are we as a profession justified in shaming the fat into losing weight, in promising them an improved state of health if they do? Why is being fat bad for you?

There are three major categories of physical risk associated with obesity: (*a*) the increased mechanical burden of obesity; (*b*) the dangers of increased food intake, or of the "wrong sort" of food; and (*c*) the miscellaneous complications and associations of obesity.

Increased Mechanical Burden

In this category are those disorders attributed to the stresses placed by extra pounds upon the skeletal frame and cardiovascular system of the fat person.

Degenerative joint disease is accelerated by obesity, especially marked in the spine, knees, ankles, and hips. This is probably mediated by the trauma of increased weight-bearing (24, 32). Many modern sports may be shown to have this same effect.

Respiratory compromise is held to be common in the fat. Certainly, dyspnea is a common symptom among the obese. This has been attributed to an increased oxygen demand (33, 34), increased intraabdominal pressure with decreased diaphragmatic activity (35), and the increased work of breathing occasioned by the mass of the fat thorax (36). Other investigators have described upper airway obstruction (37) and a decreased sensitivity to carbon dioxide (38) or oxygen (39, 40). But there is no good correlation between the presence or severity of dyspnea or alveolar hypoventilation and the body weight (39, 41). The physiologic responses of the respiratory system to weight loss are variable and ill explained (36, 42–44). The Pickwick syndrome (43) may occur in massive obesity, of course, and is probably due to a combination of mechanical and central nervous system factors. Much more work is necessary on sleep apnea in the lean (45) before any definitive statements can be made about its relationship to fatness.

Cardiomegaly occurs in the obese (41, 46). There is evidence that the cardiac enlargement of the obese is roughly proportional to the increase in body mass. This is associated with an increased blood volume and an increased cardiac output, which are also proportionate to the increased tissue mass (46, 47). In the moderately obese, this "work hypertrophy" might be the fat person's equivalent of exercise-induced cardiac enlargement, an "athlete's heart" (46). Whether the "athlete's heart" is a good or a bad thing remains controversial even in athletes (48). "Fatty heart," or degenerative intramyocardial fat droplets, are not present in the obese as a rule. There may be increased subepicardial fat and fat between myocardial muscle bundles in the very obese, but the clinical significance of this finding is uncertain (46, 49).

Surgery is more dangerous in the obese because of mechanical and anesthetic-respiratory difficulties (50, 51). There may also be a problem of delayed diagnosis of intraabdominal disease because of the increased difficulty of adequate physical examination.

Increased Food Intake

It should be noted that fat people do not continually gorge themselves. If they did, they'd perpetually gain weight. An obese individual may gain by "overeating" but, having achieved a stable weight level, will take in as many calories as are necessary to support that weight at any given level of exercise. There are data showing no marked difference in the caloric intake of a cohort of stable obese people from that of a group of normal weight (52). If a fat person exercises very little, he or she may actually eat less than a very active lean person (53).

It has been proposed that the fat are ruled by appetite—the titillation of the taste, smell, or appearance of food—more than they are by hunger (54–57). If this is true, they may be attracted to good-tasting foods such as sweets and animal fats more than the lean individual: they may have a proportionately greater intake of delicious but "unhealthy" foods.

Coronary vascular disease and its relationship to obesity has stimulated much interest, conflict, and controversy. There is evidence from the Framingham study that angina and sudden death are more common in the obese (58) with an effect that varies with age. A number of other epidemiologic studies have demonstrated little influence of obesity alone on the development of coronary heart disease (59–61). It appears that obesity alone, in the absence of hypertension or hypercholesterolemia, is not a risk factor for atherosclerotic coronary heart disease. If either or both conditions are present, however, risk is magnified by obesity (58, 62). There is a better correlation between positive caloric balance and high serum lipid levels (63) than between stable obesity and serum cholesterol levels (59, 64). Since a large number of fat people go through cyclic diets and refeeding, this potential risk of hypercholesterolemia during refeeding raises the theoretical possibility that diet therapy, with an almost invariable recidivism to obesity, may be risky.

Gallstones afflict the fat. Saturation of gallbladder bile with cholesterol (rendering it more lithogenic) is greater in the obese, and is decreased at a stable reduced weight (65). The active phase of weight reduction, however, may actually transiently increase the risk of gallstones, since fasting increases the saturation of gallbladder bile (66). Once gallstones are present, there is no evidence they are ameliorated by weight loss.

Miscellaneous Complications

Hypertension is one of the most commonly cited concomitants of obesity (67). Though some studies show only slight association of obesity with blood pressure elevation (59), the obese are probably at higher risk for hypertension than the lean. The degree of obesity is not well correlated with the severity of hypertension (41, 59), and increased pressures associate more with body weight and shape than with adiposity per se (68). Since hypertension and obesity may be genetically rather than causually related, it remains to be demonstrated that sustained weight loss leads to a sustained decrease in hypertension or its complications. Though a number of studies show diminution of blood pressure in a significant number of patients (14) when they lose weight (67, 69), the mechanism of the hypotensive effect of weight loss remains controversial, the degree variable, and the effect on prognosis obscure (3, 14).

The relationships between elevations in blood pressure and obesity obviously need further study. What we do know at present suggests that weight loss, often impossible for the patient, ought not to be a prerequisite by physicians for the institution of antihypertensive medications (70).

Diabetes mellitus is held to be a major feature of obesity. A difficulty in the interpretation of epidemiologic data is in knowing what is meant by "diabetes" in each study. There is a difference in the glucose tolerance curves of the fat compared to the lean, and this difference is lessened by weight loss (71–74). What this means is not immediately evident, since the "normals" for the standard glucose tolerance curves were established in lean young people and then applied to the obese and elderly (75). If diabetes mellitus is defined only by a glucose tolerance curve that differs from those in the lean, then many fat people are "diabetic." Yet there is no good evidence that microvascular diabetes is increased in the fat, nor that angiopathy is improved by weight loss. It is also clear that the obese differ from genetic diabetics in that circulating insulin *increases* as obesity increases (74).

Though one may argue that "decreased" glucose tolerance in the obese is normal for their weight, it would still seem prudent to advise attempts at weight reduction in those patients who have glycosuria, increased infections, cataracts, or neuropathy.

Gynecologic and obstetric influences of obesity assume added significance in that most obese people are women (76). Obese women may be at increased risk for certain complications of pregnancy, including glucose intolerance, hypertension, and thromboembolic disease (77). The detrimental effects of maternal starvation on fetal well-being are such, however, that obese pregnant women are *not* advised to lose weight during pregnancy, but rather to gain proportionate to their stage of pregnancy much as any nonobese woman would (78).

Obesity in women is associated with menstrual irregularities (79). Whether this is due to periodic dieting, the emotional stresses of obesity, or a fundamental disturbance of endocrine function in the fat woman is not yet apparent. A relationship between teenage obesity and endometrial carcinoma (80) requires further elucidation.

Hyperuricemia and gouty arthritis are increased in the obese and decline with weight loss (81, 82). In assessing hyperuricemia, especially in the obese woman, the clinician should take care to exclude thiazide diuretics (an ingredient of some diet pills) as contributory to the increased serum urate (82).

Nephrotic syndrome has been reported in four patients with massive obesity (83). What significance this has to the less-than-massively obese is uncertain.

Fatty liver may occur in fat people. A recent study (84) demonstrates fatty hepatitis, fatty fibrosis, and fatty cirrhosis in 29 overweight patients referred for study of hepatomegaly or abnormal liver functions. The question of the prevalence of fatty liver disease in the obese population not selected for evident hepatic dysfunction was not addressed. Moreover, while the 29 patients were questioned concerning potential hepatotoxic illness, alcohol, or drug abuse, none were evidently questioned about their diets. Since the obese frequently fast in attempts to lose weight, and since starvation causes fatty hepatopathy and cirrhosis, it again becomes questionable whether the obesity or its therapy was responsible for the perceived morbidity.

Psychological stresses are commented upon in many discussions of obesity (3–5, 15, 25–31). Might social attitudes concerning the obese actually create or contribute to some of its morbidity? Stress has been convincingly argued to have major influence on heart disease and cancer, cerebrovascular and circulatory disease (85). It is with these same morbid states that obesity is said to correlate. More study is called for, since, if we do stress our patients by categorizing them as diseased, we may be contributing to their disease. Perhaps our major thrust in therapy should be to change attitudes about the obese rather than their shapes.

THE VALUE OF THERAPY

In extreme, life-threatening, or immobilizing obesity, extreme measures are surely acceptable. But unnecessary dieting in mild to moderately fat people may kill them (7, 11). The multitude of complications reported from intestinal bypass surgery (6) make it inappropriate for most, if not all patients, and certainly unwarranted for cosmetic reasons alone. Addiction was (and perhaps still is) a major danger of the use of amphetamine in the therapy of obesity (86). What contribution the more potent amphetamines may have made to the increased blood pressure and possible heart disease of the obese has not been well explored.

A relatively infrequently addressed but highly significant complication of the therapy of obesity is the financial morbidity incurred by fat individuals and the nation. It is conservatively estimated that Americans spend over 10 billion dollars a year in reduction schemes of one sort or another (14). Diet foods, machines, exercises, spas, camps, clinics, pills, and corsets may drain thousands of dollars from the individual fat man or woman. What do they get for this money? Not much, it seems. Estimates are that 75–95% of all schemes for weight reduction fail (87). Even if the patients do lose weight, they rapidly regain it. And in the successful 5–25%, we really don't know what's been gained so far as health is concerned. The irony is that the therapy of obesity is generally so ineffectual that clinical studies relating weight loss to the relief of conditions in which obesity is implicated as pathogenic are nearly impossible to do (3).

It seems reasonable that one should try to treat obesity when we treat the degenerative joint disease, dyspnea, hyperglycemia, congestive heart failure, angina, or hypertension of the obese. But should we frustratingly insist on weight loss in the nonhypertensive asymptomatic obese patient?

THE ADVANTAGES OF OBESITY

The recent review of epidemiologic studies that suggested decreased mortality among the obese (13) logically raises the question of a possible selective advantage to obesity. Data exist suggesting that fat people may have better survival with cancers and tuberculosis than do the lean (14). Surprisingly, the obese also commit suicide less frequently (14). There are psychiatric testing studies which maintain that the obese may be better problem solvers and more perceptive receivers of external stimuli than the lean (55).

A well-adapted animal has the ability to store fat during periods of abundant food supply in preparation for the inevitable intervals of famine (88). It may be, then, that the prevalence of obesity in the United States is due to the conjunction of millions of years of evolutionary thrust with a remarkable sufficiency of food available at minimal exertion. The fat may be the most highly evolved among us. And should food become scarce through natural disaster, war, or shortages of energy, the fat may be the most likely survivors.

SUMMARY

In view of the conflicting data, the patent cultural biases under which we labor, and the possible harmful effects of therapy, it seems best that physicians examine the problems of obesity with the same keen scepticism and science they apply to other unsettled issues. To treat disease in the obese is obviously good. To treat simple obesity as a disease may be another matter entirely

Literature Cited

1. US Public Health Service, Division of Chronic Diseases. 1966. *Obesity and Health.* Public Health Ser. Publ. No. 1485. Washington, DC: US Govt. Print. Off.
2. Ayers, W. M. 1958. Changing attitudes toward overweight and reducing. *J. Am. Diet. Assoc.* 34:23–29
3. Mann, G. 1974. The influence of obesity on health. *N. Engl. J. Med.* 291:178–85, 226–32
4. Cahnman, W. J. 1968. The stigma of obesity. *Sociol. Q.* 9:283–99
5. Maddox, G. L., Back, K. W., Liederman, V. R. 1968. Overweight as social deviance and disability. *J. Health Soc. Behav.* 9:287–98
6. Reynolds, T. B. 1978. Medical complications of intestinal bypass surgery. *Adv. Intern. Med.* 23:47–60
7. Felig, P. 1978. Four questions about protein diets. *N. Engl. J. Med.* 298:1025–26
8. Stunkard, A. J., Rush, J. 1974. Dieting and depression reexamined. *Ann. Intern. Med.* 81:526–33
9. Rivlin, R. S. 1975. Drug therapy: therapy of obesity with hormones. *N. Engl. J. Med.* 292:26–28
10. Edison, G. R. 1971. Amphetamines: a dangerous illusion. *Ann. Intern. Med.* 74:605–10
11. Adler, A. G., Walinsky, P., Krall, R. A., Cho, S. Y. 1980. Death resulting from ipecac syrup poisoning. *J. Am. Med. Assoc.* 243:1927–29
12. Bray, G. A. 1976. The overweight patient. *Ann. Rev. Med.* 21:267–308
13. Widdowson, E. M. 1965. Chemical analysis of the body. In *Human Body Composition,* ed. J. Brozek. 7:31–47. Oxford: Pergamon
14. Bray, G. A. 1976. The obese patient. In *Major Problems in Internal Medicine,* ed. L. H. Smith, Vol. IX. Philadelphia: Saunders
15. Wilmore, J. H., Behnke, A. R. 1968. Predictability of lean body weight through anthropomorphic assessment in college men. *J. Appl. Physiol.* 5:349–55
16. Wilmore, J. H., Behnke, A. R. 1970. An anthropomorphic estimation of body density and lean body weight in young women. *Am. J. Clin. Nutr.* 23:267–74
17. Moore, F. D., Olesen, K. H., McMurrey, J. D., Parker, H. V., Ball, M. D., Boyden, C. M. 1963. *The Body Cell Mass and its Supporting Environment,* p. 484. Philadelphia: Saunders
18. Lesser, G. T., Deutsh, S., Markoffsky, J. 1971. Use of independent measurement of body fat to evaluate overweight and underweight. *Metabolism* 20:792–804
19. Society of Actuaries. 1959. *Build and Blood Pressure Study,* Vol. I, II. Chicago: Soc. Actuaries
20. Blackburn, H., Parlin, R. W. 1966. Antecedents of disease, insurance mortality experience. *Ann. NY Acad. Sci.* 134:965–1017
21. Keys, A. 1975. Overweight and the risk of heart attack. In *Obesity in Perspective. Proc. Fogarty Cent. Conf. on Obesity, Oct. 1–3, 1973.* Washington, DC: Govt. Print. Off.
22. Benn, R. T. 1970. Indices of height and weight as measures of obesity. *Br. J. Prev. Soc. Med.* 24:64
23. Andres, R. 1980. Influence of obesity on longevity in the aged. In *Aging, Cancer and Cell Membranes, Vol. VII, Advances in Cancer Biology Series.* New York: Stratton Intercontinental Medica. In press
24. Bortz, W. M. 1975. Obesity and weight control. In *Tice's Practice of Medicine.* ed. C. R. Shuman, Chap. 20, VIII:1–24. Hagerstown: Harper & Row.
25. Stunkard, A. J. 1968. Body image disturbance in obesity. In *Feelings and Their Medical Significance,* pp. 1–4. Columbus, Ohio: Ross Labs.
26. Monello, L. F., Mayer, J. 1963. Obese adolescent girls: an unrecognized "minority" group. *Am. J. Clin. Nutr.* 13:35–39
27. Richardson, S. A., Hastorf, A. H., Goodman, N., Dornbusch, S. M. 1961. Cultural uniformity in reaction to physical disabilities. *Am. Sociol. Res.* 26:241–47
28. Goodman, N., Richardson, S. A., Dornbusch, S. M., Hastorf, A. H. 1963. Variant reactions to physical disabilities. *Am. Sociol. Rev.* 28:429–35
29. Maddox, G. L., Liederman, V. 1969. Overweight as a social disability with medical implications. *J. Med. Educ.* 44:214–20
30. Maddox, G. L., Black, K. W., Liederman, V. R. 1968. Overweight as social deviance and disability. *J. Health. Soc. Behav.* 9:287–98
31. Markoutsas, E. 1979. "They said we were too fat to adopt a baby." *Good Housekeeping,* Sept. 1979, p. 96–99
32. Julkunen, H., Heinanen, O. P., Pyorala, K. 1971. Hyperostosis of the spine in an adult population: its relation to hyper-

glycemia and obesity. *Ann. Rheum. Dis.* 30:605–12

33. Mayer, J. 1966. Some aspects of the problem of regulation of food intake and obesity (concluded). *N. Engl. J. Med.* 274:722–31
34. Kaltman, A. J., Goldring, R. M. 1976. The role of circulatory congestion in the cardiorespiratory failure of obesity. *Am. J. Med.* 60:645–53
35. Lourencò, R. V. 1969. Diaphragm activity in obesity. *J. Clin. Invest.* 48:1609–14
36. Rochester, D. F., Enson, Y. 1974. Current concepts in the pathogenesis of obesity—hypoventilation syndrome. Mechanical and circulatory factors. *Am. J. Med.* 57:402–20
37. Walsh, R. E., Michaelson, E. D., Harkleroad, L. E., Zighelboim, A., Sackner, M. A. 1972. Upper airway obstruction with sleep disturbance and somnolence. *Ann. Intern. Med.* 76:185–92
38. Addington, W. W., Pfetter, S. H., Gaenster, E. A. 1969. Obesity and alveolar hypoventilation. *Respiration* 26:214–25
39. Zwillich, C. W., Sutton, F. D., Pierson, D. J., Creagh, E. M., Weil, J. V. 1975. Decreased hypoxic ventilatory drive in the obesity-hypoventilation syndrome. *Am. J. Med.* 59:343–48
40. Kronenberg, R. S., Drage, C. W., Stevenson, J. E. 1977. Acute respiratory failure and obesity with normal ventilatory response to carbon dioxide and absent hypoxic ventilatory drive. *Am. J. Med.* 62:772–76
41. Alexander, J., Ahmad, K., Cole, V. 1962. Observations on some clinical features of extreme obesity, with particular reference to cardiopulmonary respiratory effects. *Am. J. Med.* 32:512–24
42. Emirgil, C., Sobol, B. J. 1973. The effects of weight reduction on pulmonary function and the sensitivity of the respiratory center in obesity. *Am. Rev. Respir. Dis.* 108:831–42
43. Burwell, C. S., Robin, E. D., Whaley, R. D., Bickelman, A. G. 1956. Extreme obesity associated with alveolar hypoventilation. A Pickwickian syndrome. *Am. J. Med.* 21:811–18
44. Fried, P. I., McClean, P. A., Phillipson, E. A., Zamel, N., Murray, F. T., Marliss, E. B. 1976. Effect of ketosis on respiratory sensitivity to carbon dioxide in obesity. *N. Engl. J. Med.* 204:1081–86
45. Guilleminault, C., Tilkian, A., Dement, W. C. 1976. The sleep apnea syndromes. *Ann. Rev. Med.* 27:465–84
46. Amad, K. H., Brennan, J. C., Alexander, J. K. 1965. The cardiac pathology of chronic exogenous obesity. *Circulation* 32:740–45
47. Kaltman, A. J., Goldring, R. M. 1976. Role of circulatory congestion in the cardiorespiratory failure of obesity. *Am. J. Med.* 60:645–53
48. Crawford, M. H., O'Rourke, R. A. 1979. The athlete's heart. *Adv. Intern. Med.* 23:311–29
49. Cryer, P. E., Kissane, J. M. 1978. Massive obesity and cardiac failure. *Am. J. Med.* 64:827–33
50. Warner, W. A. 1968. The obese patient and anesthesia. *J. Am. Med. Assoc.* 205:102–3
51. Keys, A. 1955. Relative obesity and its health significance. *Diabetes* 4:447–55
52. Thomson, A. M., Billewicz, N. Z., Passmore, R. 1961. The relationship between caloric intake and body weight in man. *Lancet* 1:1027–28
53. Nelson, R. A., Anderson, L. F., Gastineau, C. F., Hayles, A. B., Stamnes, C. L. 1973. Physiology and natural history of obesity. *J. Am. Med. Assoc.* 223:637–30
54. Schachter, S. 1968. Obesity and eating. *Science* 161:751–56
55. Schachter, S. 1971. Some extraordinary facts about obese humans and rats. *Am. Psychol.* 26:129–44
56. Nisbett, R. E. 1968. Determinants of food intake in obesity. *Science* 159:1254–55
57. Stunkard, A. J., Fox, S. 1971. The relationship of gastric motility and hunger: a summary of the evidence. *Psychosom. Med.* 33:123–34
58. Kannel, W. B., LeBauer, E. J., Dawber, T. R., McNamara, P. M. 1967. Relation of body weight to development of coronary heart disease. *Circulation* 35: 734–44
59. Weinsier, R. L., Fuchs, R. S., Kay, T. D., Triebwasser, J. H., Lancaster, M. C. 1976. Body fat, its relationship to coronary heart disease, blood pressure, lipids, and other risk factors measured in a large male population. *Am. J. Med.* 61:815–24
60. Chapman, J. M., Coulson, A. H., Clark, V. A., Borun, E. R. 1971. The differential effect of serum cholesterol, blood pressure, and weight on the incidence of myocardial infarction and angina pectoris. *J. Chron. Dis.* 23:631–45
61. Rosenman, R. H., Friedman, M., Straus, R., Jenkins, C. D., Zyzanski, S. J., Wurm, M. 1970. Coronary heart disease in the western collaborative group study: a followup experience of 4½ years. *J. Chron. Dis.* 23:173–90

62. Keys, A., Aravanis, C., Blackburn, H., Van Buckem, F. S. P., Busina, R., Djordevic, B. S., Findanza, F. F., Karvonen, M. J., Menolti, A., Puddu, V., Taylor, H. L. 1972. Coronary heart disease: overweight and obesity as risk factors. *Ann. Intern. Med.* 77:15–27
63. Walker, W. J., Lawry, E. Y., Love, D. E., Levine, S. A., Stare, F. J. 1953. Effect of weight reduction and caloric balance on serum lipoprotein and cholesterol levels. *Am. J. Med.* 14:654–64
64. Montoye, H. J., Epstein, F. H., Kjelsberg, M. O. 1966. Relationship between serum cholesterol and body fatness. *Am. J. Clin. Nutr.* 18:397–406
65. Bennion, L. J., Grundy, S. M. 1975. Effects of obesity and caloric intake on biliary lipid metabolism in man. *J. Clin. Invest.* 56:996–1011
66. Goldstein, L. L., Schoenfield, L. J. 1975. Gallstones: pathogenesis and medical treatment. *Adv. Intern. Med.* 20:89–118
67. Chiang, B. N., Perlman, L. V., Epstein, F. H. 1969. Overweight and hypertension: a review. *Circulation* 39:403–21
68. Robinson, S. C., Brucer, M. 1940. Hypertension: Body build and obesity. *Am. J. Med. Sci.* 199:819–29
69. Reisin, E., Abel, R., Modan, M., Silverberg, D. S., Eliahin, H. E., Modan, B. 1978. Effect of weight loss without salt restriction on the reduction of blood pressure in overweight hypertensive patients. *N. Engl. J. Med.* 298:1–6
70. Tobian, L. 1978. Hypertension and obesity. *N. Engl. J. Med.* 298:46–48
71. Bortz, W. M. 1969. Metabolic consequences of obesity. *Ann. Intern. Med.* 71:833–43
72. Felig, P., Marliss, E., Cahill, G. F. 1969. Plasma amino acid levels and insulin secretion in obesity. *N. Engl. J. Med.* 281:811–16
73. Kreisberg, R. A., Boshell, B. R., DiPlacido, J., Roddam, R. F., 1967. Insulin secretion in obesity. *N. Engl. J. Med.* 276:314–19
74. El-Khodary, A. Z., Ball, M. F., Oweiss, I. M., Canary, J. J. 1972. Insulin secretion and body composition in obesity. *Metabolism* 21:641–55
75. Siperstein, M.D. 1975. The glucose tolerance test: A pittfall in the diagnosis of diabetes mellitus. *Adv. Intern. Med.* 20:297–323
76. *Health, United States, 1978: 3rd Ann. Rep. on Health Status of the Nation,* Dept. HEW Publ. No. (PHS) 78-1232. Hyattsville, Md: US Govt. Print. Off.
77. Peckham, C. H., Christianson, R. E. 1971. The relationship between prepregnancy weight and certain obstetric factors. *Am. J. Obstet. Gynecol.* 11:1–7
78. Pitkin, R. M. 1977. Nutritional influences during pregnancy. *Med. Clin. North Am.* 61:3–15
79. Hartz, A., Wong, A., Katayama, K. P., Rimm, A. A. 1979. The association of obesity with anovulatory cycles and related menstrual irregularities in 36,081 women. *Int. J. Obesity* 3:57–73
80. Blitzer, P. H., Blitzer, E. C., Rimm, A. A. 1976. Association between teenage obesity and cancer in 56,111 women. *Prev. Med.* 5:20–31
81. Rimm, A. A., Werner, L. H., VanYserloo, B., Bernstein, R. A. 1975. Relationship of obesity and disease in 73,532 weight-conscious women. *Public Health Rep.* 90:44–51
82. Wyngaarden, J. B. 1974. Gout and other disorders of uric acid metabolism. In *Harrison's Principles of Internal Medicine,* ed. M. M. Wintrobe, G. W. Thorn, R. D. Adams, E. Braunwald, K. I. Isselbacher, R. G. Petersdorf, Edition VII, pp. 607–17. New York: McGraw-Hill
83. Weisinger, J. R., Kempson, R. L., Eldridge, F. L., Swenson, R. S. 1974. The nephrotic syndrome: a complication of massive obesity. *Ann. Intern. Med.* 81:440–47
84. Adler, M., Schaffner, F. 1979. Fatty liver hepatitis and cirrhosis in obese patients. *Am. J. Med.* 67:811–16
85. Hurst, M. W., Jenkins, C. D., Rose, R. M. 1976. The relation of psychological stress to onset of medical illness. *Ann. Rev. Med.* 27:301–12
86. United States, Senate Committee on the Judiciary, Subcommittee on Antitrust and Monopoly. 1968. Hearings on the diet pill industry. 90th Congress, 2nd session
87. Bray, G. A., ed. 1979. *Obesity in America,* NIH Publ. No. 79-359. US Dept. HEW Public Health Serv., Natl. Inst. Health
88. Neel, J. F. 1962. Diabetes mellitus: a "thrifty" genotype rendered detrimental by "progress"? *Am. J. Hum. Genet.* 14:353–62

Ann. Rev. Med. 1981. 32:233–43

TOTAL PARENTERAL NUTRITION IN THE MANAGEMENT OF THE CANCER PATIENT[1]

❖7395

Murray F. Brennan, M.D.

Surgical Metabolism Section, Surgery Branch, National Cancer Institute,
National Institutes of Health, Bethesda, Maryland 20205

Introduction

Anorexia and diminished oral intake have long been known to accompany cancer (1). If anorexia and diminished intake were the sole cause of the weight loss, weakness, tissue depletion, and biological bankruptcy that is embodied in the term cachexia, then the advent of intravenous feeding or total parenteral nutrition (TPN) would have been a therapeutic triumph for the cancer patient. Unfortunately, the degree of cachexia seems independent of oral intake, tumor burden, cell type, or anatomical site of the neoplasm. The anorexia and weight loss are simply explained if the neoplasm involves the oropharynx or esophagus and thereby prevents adequate intake. However, many neoplasms not involving the gastrointestinal tract are similarly accompanied by major degrees of inanition.

If, then, the inability to ingest, digest, or absorb is not the sole cause of the malnutrition that accompanies cancer, then the widespread use of TPN in the cancer patient must be justified by its effectiveness in reversing nutritional morbidity. Although rigorous clinical studies in man have not always been performed, there is evidence in newborns and animals that TPN can support growth and development. It seems reasonable to assume then that TPN is an effective treatment for uncomplicated malnutrition due to inadequate intake (2, 3).

Efficacy in Nonmalignant Disease

Generally accepted conditions thought to require nutritional support are listed in Table 1. The widespread use of total parenteral nutrition as a disease-oriented therapeutic modality rests on insecure ground. Rigorous evidence to support indiscriminate use of TPN for all of these indications is lacking. Mortality from congenital and acquired gastrointestinal abnormalities, and intractable diarrhea in infants who cannot be maintained by the enteral route, diminished from 75% to 10% with the introduction of TPN; it is reasonable to conclude that TPN played a major role in this decrease (4, 5).

---

[1] The US Government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this paper.

Dietary management of inborn errors of metabolism has prevented early death or severe mental retardation, so that in situations where enteral support has been impractical TPN can probably be similarly efficacious (6, 7). Reasonable data exist to support the use of TPN in adolescents with inflammatory bowel disease, where increased growth rate and growth spurts have been demonstrated (8, 9).

Information supporting the use of TPN in the low birth weight infant has been seriously questioned. But formerly 50–70% of such infants had severe developmental abnormalities (10), and only 15% are now abnormal (11). Survival of infants weighing 1000–1500 g has improved, from 40% to near 80%, in association with the improved intensive neonatal support provided from 1953 to 1969 (11). The severely burned patient and the patient with established gastrointestinal fistulae not due to cancer also benefit from TPN (3, 12).

In patients with acute renal failure, prospective randomized studies have been performed in a relatively homogenous patient group (14). In one study, although survival from renal failure was significantly improved in the TPN group compared to the glucose control group, no decrease in need for dialysis was identified and in-hospital mortality was not significantly improved (14). In a second study, no improvement in survival or in duration of renal failure was noted in the TPN group (15).

In malnourished patients undergoing cardiac surgery, a randomized trial of immediate (five-day) postoperative TPN vs no intravenous nutritional support showed no noticeable effect on morbidity or mortality (16). The data for modified TPN mixtures rich in the branched-chain amino acids is less secure, but suggestive of a positive effect (17–19).

Table 1 Conventionally accepted indications for total parenteral nutrition

1. Congenital or acquired malfunction of the gastrointestinal tract
    A. infants and the newborn
    B. adults
2. Unremitting diarrhea of infancy
3. Inborn errors of metabolism
4. Low birth weight infants
5. Inflammatory bowel disease
    A. children and adolescents
    B. adults
6. Renal failure
7. Hepatic failure
8. Cardiopulmonary failure
9. Severely burned hypercatabolic man

Generalized agreement on the need for parenteral nutritional support in hypercatabolic man unable to ingest has recently been articulated, but is qualified by comments about the scientific inadequacy of the supportive data (3).

An estimated 1 to 2% of all in-hospital patients are supported by TPN. When one considers that the average approximate cost to the patient is $60 per 1000 cc, then the cost of one day's TPN is approximately the average room cost. If TPN saves one in-hospital day it is cost effective! Unfortunately, data on the usage and cost of TPN nationwide is currently unavailable. One estimate suggests that 150,000–200,000 patients per year receive TPN; one third of these have cancer (J. L. Mullen, personal communication).

Indications in the Cancer Patient

Indications for nutritional support are relatively easy to categorize, despite a lack of meaningful data as to the significance of such indications in the cancer-bearing host. Any patient whose hospital stay is expected to exceed 15 days, who is unable to ingest, absorb, or digest nutrient via the gastrointestinal tract is a potential candidate for intravenous nutritional support.

The peripheral determinations of anthropometrics, serum albumin, serum proteins, and acute-phase reactants such as transferrin can help support such a decision and, in a few instances, serve to monitor nutritional progress (20). Unfortunately, a prospective search for factors that would predict successful nutritional repletion in patients with advanced malignancy has been relatively unrewarding (21).

The more difficult decision, however, is whether or not the patient has realistic prospects for meaningful, antineoplastic therapy. This question may be irrelevant, however, unless we can show that intravenous nutritional support can reverse the malnutrition and metabolic defects that accompany cancer.

Benefits in the Cancer Patient

SUPPRESSION OF GLUCONEOGENESIS The cancer-bearing host, like any starving patient, depends to some degree on gluconeogenesis from amino acids for maintenance of blood glucose. This continuous strain on the body protein stores inevitably leads to lean tissue mass dissolution and death from starvation. Several authors (22–25) have shown that intravenous glucose or TPN mixtures can suppress gluconeogenesis from amino acid precursors in the cancer-bearing host.

REVERSAL OF WEIGHT LOSS Many studies demonstrate that weight gain can be achieved on total parenteral nutrition in cancer (26–28). More recently, however, considerable concern has been expressed as to whether or not this weight gain reflects fat accretion, lean tissue accretion, or some mixture of both. Some observers suggest that fat repletion is much more readily obtained than is lean tissue repletion (29). In some patients who have responded to their antineoplastic therapy, lean tissue can be accrued, as measured by total body potassium determinations or by nitrogen balance, although at a cost of increased amino nitrogen flux (25).

PROVISION OF EXCESS ENERGY NEEDS Some evidence now exists that the cancer-bearing patient is less able to adapt to the rigors of starvation than his non-cancer-bearing, similarly malnourished control (30, 31). The evidence that the cancer-bearing host has increased energy demands is somewhat confusing (32, 33); recent evidence suggests that patients with widely metastatic cancer do indeed have increased energy needs (34). Where increased need for calories exists it can be met, provided liver function is carefully monitored.

REVERSAL OF DEFICIENCY STATES Deficiencies of electrolytes, minerals, trace metals, and vitamins are commonly implicated in both causation and progression of malignancy. Few data exist to support or refute these suggestions. Total parenteral nutrition can readily reverse all of these abnormalities, and no patient should suffer from such deficiency.

REVERSAL OF IMMUNE DEFICIENCIES The cancer-bearing host is commonly immunosuppressed. This immunosuppression may be due to the patient's disease or associated infectious complications, or it may be a product of immunosuppressive therapy. Impressive claims have been made that this immunosuppression can be reversed by intravenous feeding (35); however, whether the concomitant reversal of the patient's relative anergy is a response to antineoplastic therapy or to the associated nutritional support is unclear. Such elucidation awaits difficult-to-perform clinical studies. The preliminary results in animals do suggest an independent role for nutrition in the reversal of such immune defects (36).

Technical Considerations for TPN

VENOUS ACCESS Venous access is a recurring problem because of the chronic peripheral venous sclerosis that accompanies the multiple venesections necessary to administer sclerotic chemotherapeutic agents or to sample venous blood for hematological monitoring. The problem is compounded by the desire to avoid placing large-bore needles into central veins when platelet and white cell counts are low. Various methods have been employed to improve long-term access to such patients. The most popular are the Broviac and Hickman long-term, in-dwelling, silastic catheters. The placement of these catheters requires a formal surgical procedure. Central silastic catheters are less irritating than the conventional, widely-used polyvinylchloride catheters (37), and there is a high incidence of superior vena caval thrombosis (25% in an autopsy study). This makes the use of a percutaneous, silicon catheter much more appealing (38). Such a catheter does exist (VICRA—Division of Travenol Laboratories, Dallas, TX 75220). It can be inserted either centrally (Centrasil®) or peripherally (Intrasil®). Unfortunately, the peripheral route is rarely available and the central catheter is not widely accepted (39). Prospective studies are not available comparing the two different types of catheter in cancer patients. Similarly, the true incidence of superior vena cava or subclavian vein thrombosis is not known. The high incidence found at autopsy occurred at a time when techniques for total parenteral nutrition were still evolving (38). More recently, a review of controlled studies done by the Diet, Nutrition, and Cancer Program of the National Cancer Institute suggested a clinically detectable incidence of approximately 5% (40). Recently, as part of a prospective randomized study of TPN, a 10% incidence of subclavian vein thrombosis in 33 patients, with an additional 33% incidence of phlebitis, (8 of 13 requiring catheter removal) was reported (41). It is my own belief that the true incidence of subclavian venous thrombosis is much higher and closer to that of the autopsy series than is readily accepted. Because subclavian venous thrombosis is often of slow onset without evidence of venous obstruction, the clinical significance is less important. Reports of death or complications resulting from subclavian venous thrombosis are few, and the patient who slowly thromboses his subclavian vein may rarely, if ever, be at risk. Only in patients requiring repeated access to the central venous route is such a problem manifest.

Some chemotherapeutic agents are compatible with TPN mixtures, which suggests that such agents will be much more frequently instilled centrally, with an assumed increase in thrombotic complications (42).

FEVER Fever commonly accompanies long-term parenteral nutrition in the cancer-bearing host, particularly during periods of immunosuppression and granulocyte depletion. We have elected in the presence of a fever not to routinely remove what is often the patient's only source of venous access. We have become more conservative, relying on the appearance of new glucose intolerance or positive blood cultures before removal of the line. The catheter can be readily exchanged over a guide wire with minimal discomfort or risk to the patient and without losing a valued access line (43). On rare occasions when no other options are available, the treatment of a fever or a positive blood culture can be performed without removing the line. It should be strongly emphasized that where other access routes are available, the safe and prudent course is the removal of the catheter (for unexplained fever), with subsequent replacement once the fever has resolved.

METABOLISM In any well-run total parenteral nutrition service, the majority of metabolic complications should be no greater than the incidence seen in patients with similar degrees of malignancy not receiving TPN (40). Clear indications have now been provided for the delivery of vitamins; the failure to support patients with a trace metal preparation will result in inevitable serum depletion and, over the long term, the clinical syndromes of zinc and copper deficiency (44–48). The relative importance of selenium and chromium deficiencies are now being elucidated. Plasma deficiencies of choline have been identified, and whether or not TPN solutions should be supplemented with choline awaits the results of a prospective randomized study (48a). Essential fatty acid deficiency will uniformly develop if essential fatty acids are not provided (47, 49), with 1000 cc of fat emulsion per week being adequate for prevention.

The current concerns about possible metabolic defects accompanying excess glucose loads may mean that use of intravenous fat as a caloric source rather than as an essential fatty acid source will become more widespread (50). Given the wide disparity in cost, this will depend upon proven efficacy. It is well appreciated that high-dose glucose infusion can, on occasions, precipitate lactic acidosis in the cancer-bearing host (51). Of great interest are some recent animal studies suggesting that a TPN mixture high in fat calories showed the best relative host preservation without tumor stimulation, of the mixtures employed (52).

The current concerns about the use of intravenous iron dextran seem to be overemphasized, and we have begun to add iron to the parenteral nutritive mixture in an effort to determine iron requirements (53). It is clear that with current aggressive blood, plasma, and serum value monitoring, the patient can easily lose by venesection a unit of blood every 2–4 weeks. This is a loss to the body iron stores of approximately 200 mg.

TUMOR STIMULATION The question of whether or not nutritional support is really "feeding the tumor" must be raised. Here the data are virtually nonexistent. In animals, host preservation without marked tumor growth stimulation can be demonstrated in short-term studies of intravenous nutrition when compared to anorectic cancer-bearing controls (54–56).

A recent observation that long-term animal TPN resulted in host maintenance and increased tumor size but no prolongation of survival awaits confirmation (M. B. Popp, S. D. Morrison, M. F. Brennan, unpublished observations). The question of nutrition stimulating tumor growth is inadvertently addressed each time a cancer patient on TPN receives what proves subsequently to be ineffective anticancer therapy. Indeed, in one randomized study, patients with metastatic colorectal cancer receiving ineffective chemotherapy were randomized to TPN or no TPN: The survival time was shorter in the TPN-supported group (57), though stimulation of measurable tumor burden was not identified as the cause of early death (57). In a panel report presented by the NIH, no documented tumor stimulation in man was reported (58).

Clinical Trials

There are innumerable reports attesting to the benefits of total parenteral nutrition in the cancer-bearing host (26–28, 59). All of these studies suggest that intravenous nutritional support can be life-saving in some patients undergoing aggressive therapy for cancer.

The more difficult question of the exact role that nutritional support should play in the management of the cancer patient is only now being examined in a critical manner. The initial study, in which only those patients with greater than 6 percent body weight loss who received TPN responded to their chemotherapy for non-oat-cell carcinoma of the lung, received tremendous emphasis (60). This retrospective study highlighted some of the problems in the evaluation of response and the role of nutritional support in that response. Since that time, a number of prospective randomized studies have been initiated and are beginning to be reported (Table 2). The results have certainly not been dramatic in terms of tolerance to therapy whether radiation, chemotherapy, surgery, or combinations of all three, no decrease in the nadir of the white count or platelet having been demonstrated (66, 68) except in the granulocyte series in one study (64). Operative deaths were not different in two studies (61, 62, 65) nor was tolerance to x-ray therapy (63) or chemotherapy (66). The metabolic efficacy of the TPN employed highlights some of the problems mentioned above. Although weight gain occurred in some patients when compared to controls who did not receive TPN (41, 61, 62, 64), others showed diminution in weight loss (28, 41, 69, 70). Decrease in complications of therapy was seen in only one study (70).

Summary and Conclusions

There can no longer be any doubt that selected patients, for whom effective antineoplastic therapy is either available or has been given, can be prevented from dying from malnutrition due to inadequate oral intake by the judicious use of TPN. Total parenteral nutrition can now be a safe, practical, and most efficacious way of reversing malnutrition. Extrapolation from these dramatic examples to all patients with cancer who have minimal malnutrition or ineffective neoplastic therapies is not justified. The exact indications for the use of nutritional support by the intravenous route in the patient with cancer are still being evaluated. TPN can, like saline, be life saving (71). We should, in 25 years, achieve the critical appraisal of TPN that it has taken 150 years to accomplish for intravenous saline (72).

Table 2 TPN in the management of cancer: randomized trials and clinical efficacy

Reference		Number of patients	Therapy[a]	Site type	CR[b]	PR	Survival median (weeks)
61	TPN	30	S	G.I.	—	—	—
62	No TPN	26		adenoca.	—	—	—
63	TPN	11	XR	pelvic,	—	45%	—
	No TPN	9		abdom. ca	—	33%	—
64	TPN	13	C–I	lung	0	31%	—
	No TPN	13		squamous	0	7%	—[c]
28	TPN	21	XR	abdom	—	—	7
	No TPN	19		pelvic ca	—	—	1
28	TPN	42	XR	advanced	—	—	33
	No TPN	39		ovarian	—	—	36
65	TPN	38	S	esophagus,	—	—	—
	No TPN	36		gastric ca	—	—	—
66	TPN	15	C	DHL[d]	—	—	—
	No TPN	18			—	—[c]	—[c]
57	TPN	20	C	metastatic	—	—	11
	No TPN	25		colorectal	—	—	43
41	TPN	13	C	oat-cell	85%	15%	—
	No TPN	17			59%	41%	—
67	TPN	21	C	lung	—[c]	—[c]	—
	No TPN	21		oat-cell			—[c]
68	TPN	14	C	lung	—	14%	11
	No TPN	13		non-oat-cell	—	23%	12

[a] S = surgery, C = chemotherapy, I = immunotherapy, XR = Radiation therapy.
[b] Not available.
[c] No difference TPN vs No TPN.
[d] DHL = diffuse histiocytic lymphoma.

Literature Cited

1. Costa, G. 1977. Cachexia the metabolic component of neoplastic diseases. *Cancer Res.* 37:2327–35
2. Dudrick, S. J., Wilmore, D. W., Vars, H. M., Rhoads, J. E. 1968. Long-term parenteral nutrition with growth, development, and positive nitrogen balance. *Surgery* 64:134–42
3. Wilmore, D. W., Kinney, J. M. 1980. Report from the Panel on Trauma and Infection. *Proc. NIH Conf. on Nutritional Support of the Patient,* ed. R. Edelman, C. W. Callaway. *Am. J. Clin. Nutr.* In press
4. Filler, R. M., Eraklis, A. J., Rubin, V. G., Das, J. B. 1969. Long-term parenteral nutrition in infants. *N. Engl. J. Med.* 281:589–91
5. Keating, J. P., Ternberg, J. L. 1971. Amino acid-hypertonic glucose treatment for intractable diarrhea in infants. *Am. J. Dis. Child.* 122:226–28
6. Baumeister, A. A. 1967. The effects of dietary control on intelligence in phenylketonuria. *Am. J. Ment. Def.* 71: 840–47
7. Snyderman, S. E., Norton, P. M., Roitman, E., Holt, L. E. Jr. 1964. Maple syrup urine disease, with particular reference to dietotherapy. *Pediatrics* 34:454–72
8. Grand, R. J., Shen, G., Werlin, S. L., Kelts, D. G., Boehm, C. 1977. Reversal of growth arrest in Chron's disease: A new approach. *Pediatr. Res.* 11:444 (Abstr.)
9. Werlin, S. L., Grand, R. J. 1977. Severe colitis in children and adolescents: diagnosis, course and treatment. *Gastroenterology* 73:828–32
10. Lubchenco, L. O., Horner, F. A., Reed, L. H., Hix, I. E., Metcalf, D., Cohig, R., Elliott, H. C., Bourg, M. 1963. Sequelae of premature birth. *Am. J. Dis. Child.* 106:101–15
11. Rawlings, G., Reynolds, E. O. R., Stewart, A., Stang, L. B. 1971. Changing prognosis for infants of very low birth weight. *Lancet* 1:516–19
12. Jeejeebhoy, K. N., Rosenberg, I. H. 1980. Report from the Panel on Gastrointestinal Diseases. See Ref. 3
13. Deleted in proof
14. Abel, R. M., Beck, C. H., Abbott, W. M., Ryan, J. A., Barnett, G. O. 1973. Improved survival from acute renal failure after treatment with intravenous essential L-amino acids and glucose. *N. Engl. J. Med* 288:695–99
15. Leonard, C. D., Luke, R. G., Siegel, R. R. 1975. Parenteral essential amino acids in acute renal failure. *Urology* 6(2):154–57
16. Abel, R. M., Fischer, J. E., Buckley, M. J., Barnett, G. O., Austen, W. G. 1976. Malnutrition in cardiac surgical patients: results of a prospective randomized evaluation of early postoperative parenteral nutrition. *Arch. Surg.* 111:45–50
17. Fischer, J. E., Funovics, J. M., Aguirre, A., James, J. H., Keane, J. M., Wesdorp, R. I. C., Yoshimura, N., Westman, T. 1975. The role of plasma amino acids in hepatic encephalopathy. *Surgery* 78:276–90
18. Fischer, J. E., Rosen, H. M., Eberd, A. M., James, J. H., Keane, J. M., Soeters, P. B. 1976. The effect of normalization of plasma amino acids on hepatic encephalopathy in man. *Surgery* 80:77–91
19. Fischer, J. E., Fishman, A. 1980. Report from the Panel on Liver, cardiopulmonary and renal failure. See Ref. 3
20. Bistrian, B. R., Blackburn, G. L., Vitale, J., Cochran, D. 1976. Prevalence of malnutrition in general medical patients. *J. Am. Med. Assoc.* 235:1567–70
21. Ericksson, B., Douglass, H. O. 1980. Intravenous hyperalimentation: an adjunct to treatment of malignant disease of upper gastrointestinal tract. *J. Am. Med. Assoc.* 243:2049–52
22. Waterhouse, C., Kemperman, J. H. 1971. Carbohydrate metabolism in subjects with cancer. *Cancer Res.* 31: 1273–78
23. Waterhouse, C., Jeanpetre, N., Keilson, J. 1979. Gluconeogenesis from alanine in patients with progressive malignant disease. *Cancer Res.* 39:1968–72
24. Waterhouse, C. 1980. Oxidation and metabolic interconversions in malignant cachexia. *Cancer Treat. Rep.* In press
25. Brennan, M. F., Burt, M. E. 1980. Nitrogen metabolism in cancer. *Cancer Treat. Rep.* In press
26. Copeland, E. M., MacFadyen, B. V., Lanzotti, V., Dudrick, S. J. 1975. Intravenous hyperalimentation as an adjunct to cancer chemotherapy. *Am. J. Surg.* 129:167–70
27. Copeland, E. M., Daly, J. M., Dudrick, S. J. 1977. Nutrition as an adjunct to cancer treatment in the adult. *Cancer Res.* 37:2451–56
28. Solassol, C., Joyeuz, H., DuBois, J. B. 1979. Total parenteral nutrition (TPN) with complete nutritive mixtures. An artificial gut in cancer patients. *Nutr. Cancer* 1:13–18

29. Nixon, D., Rudman, D., Heymsfield, S., Ansley, J., Kutner, M. 1978. Response to nutritional support in cachetic patients. *Am. Assoc. Cancer Res.* #698 (Abstr.)
30. Brennan, M. F. 1977. Uncomplicated starvation versus cancer cachexia. *Cancer Res.* 37:2359–64
31. Warnold, I., Falkheden, T., Hulten, B., Isaakson, B. 1978. Energy intake and expenditure in selected groups of hospital patients. *Am. J. Clin. Nutr.* 31: 742–49
32. Warnold, I., Lundholm, K., Schersten, T. 1978. Energy balance and body composition in cancer patients. *Cancer Res.* 38:1801
33. Young, V. R. 1977. Energy metabolism and requirements in the cancer patient. *Cancer Res.* 37:2336–47
34. Steinberg, J. L., Crosby, L. O., Feurer, I. D., Mullen, J. L. 1980. Indirect calorimetry and cancer patients. *Am. Soc. Clin. Oncol.* C-229 (Abstr.)
35. Copeland, E. M., MacFadyen, B. V., Dudrick, S. J. 1976. Effect of intravenous hyperalimentation on established delayed hypersensitivity in the cancer patient. *Ann. Surg.* 184:60–64
36. Daly, J. M., Copeland, E. M., Dudrick, S. J. 1978. Effect of intravenous nutrition on tumor growth and host immunocompetence in malnourished animals. *Surgery* 84:655–58
37. Welch, G. W., McKeel, D. W., Silverstein, P., Walker, H. L. 1974. The role of catheter composition in the development of thrombophlebitis. *Surg. Gynecol. Obstet.* 138:421–24
38. Ryan, J. A., Abel, R. M., Abbott, W. M., Fischer, J. E. 1974. Catheter complications of total parenteral nutrition. *N. Engl. J. Med.* 290:757–61
39. Abrahm, J., Mullen, J. L., Jacobson, N., Polomano, R. 1980. Chronic central venous access in patients with leukemia. *Cancer Treat. Rep.* In press
40. Mullen, J. L. 1980. Complications of total parenteral nutrition in cancer patients. *Cancer Treat. Rep.* In press
41. Valdivieso, M., Bodey, G. P., Benjamin, R. S., Barkley, H. T., Freeman, M. B., Ertel, M., Smith, T. L., Mountain, C. F. 1980. Role of intravenous hyperalimentation as an adjunct to intensive therapy for small cell bronchogenic carcinoma: preliminary observations. *Cancer Treat. Rep.* In press
42. Hardin, T., Clibon, U., Page, C., Aust, J. B. 1980. Compatibility of 5-fluorouracil with total parenteral nutrition solution in glass and plastic containers. *Am. Soc. Clin. Oncol.* C-71 (Abstr.)
43. Maher, M. M., Brennan, M. F. 1980. Central venous catheter exchange using a guide wire. *Proc. Am. Soc. Parenteral and Enteral Nutrition* Nu4 (Abstr.)
44. Nichoalds, G. E., Meng, H. C., Caldwell, M. D. 1977. Vitamin requirements in patients receiving total parenteral nutrition. *Arch. Surg.* 112:1061–64
45. Lowry, S. F., Goodgame, J. T., Maher, M. M., Brennan, M. F. 1978. Parenteral vitamin requirements during intravenous feeding. *Am. J. Clin. Nutr.* 31: 2149–58
46. Lowry, S. F., Goodgame, J. T., Smith, J. C., Maher, M. M., Makuch, R. W., Henkion, R. I., Brennan, M. F. 1979. Abnormalities of zinc and copper during total parenteral nutrition. *Ann. Surg.* 189:120–28
47. McCarthy, D. M., May R. J., Maher, M., Brennan, M. F. 1978. Trace metal and essential fatty acid deficiency during total parenteral nutrition. *Am. J. Dig. Dis.* 23:1009–16
48. Lowry, S. F., Smith, J., Brennan, M. F. 1980. Zinc and copper replacement during total parenteral nutrition. *Am. J. Clin. Nutr.* In press

48a. Burt, M. E., Hanin, I., Brennan, M. F. 1980. Choline deficency in man: association with total parenteral nutrition. *Lancet* 2:638–39

49. Goodgame, J. T., Lowry, S. F., Brennan, M. F. 1978. Essential fatty acid deficiency in total parenteral nutrition: time course of development and suggestions for therapy. *Surgery* 84:271–77
50. Askanazi, J., Elwyn, D. H., Silverberg, P. A., Rosenbaum, S. H. Kinney, J. M. 1980. Respiratory distress secondary to a high carbohydrate load: a case report. *Surgery* 87:596–98
51. Goodgame, J. T., Pizzo, P., Brennan, M. F. 1978. Iatrogenic lactic acidosis: Association with hypertonic glucose administration in a patient with cancer. *Cancer* 42:800–3
52. Buzby, G. P., Mullen, J. L., Stein, T. P., Miller, E. E., Hobbs, C. L., Rosato, E. F. 1980. Host-tumor interaction and nutrient supply. *Cancer* 45:2940–48
53. Hamstra, R. D., Block, M. H., Schocket, A. C. 1980. Intravenous iron dextran in clinical medicine. *J. Am. Med. Assoc.* 243:1724–31
54. Goodgame, J. T., Lowry, S. F., Brennan, M. F. 1979. Nutritional manipulations and tumor growth. II. The effects of intravenous feeding. *Am. J. Clin. Nutr.* 32:2285–94

55. Lowry, S. F., Goodgame, J. T., Norton, J. A., Jones, D. C., Brennan, M. F. 1979. Effect of chronic protein malnutrition on host-tumor composition and growth. *J. Surg. Res.* 26:79–86
56. Steiger, F., Oram-Smith, J., Miller, E., Kuo, L., Vars, H. M. 1975. Effects of nutrition on tumor growth and tolerance to chemotherapy. *J. Surg. Res.* 18:455–61
57. Nixon, D., Kutner, M. H., Ansley, J., Lawson, D. H., Lynn, M. J., Rudman, D. 1980. Survival with and without hyperalimentation. *Cancer Treat. Rep.* In press
58. Brennan, M. F., Copeland, E. M. 1980. Report from the Panel on Cancer. See Ref. 3
59. Harvey, K. B., Bothe, A., Blackburn, G. L. 1979. Nutritional assessment and patient outcome during oncological therapy. *Cancer* 43:2065–69
60. Lanzotti, V. C., Copeland, E. M., George, S. L., Dudrick, S. J., Samuels, M. I. 1975. Cancer chemotherapeutic response and intravenous hyperalimentation. *Cancer Chemother. Rep.* 59: 437–39
61. Holter, A. R., Rosen, H. M., Fischer, J. E. 1976. The effects of hyperalimentation on major surgery in patients with malignant disease a prospective study. *Acta Chir. Scand.* (*Suppl.*) 466:86–87
62. Holter, A. R., Fischer, J. E. 1977. The effects of perioperative hyperalimentation on complications in patients with carcinoma and weight loss. *J. Surg. Res.* 23:31–34
63. Valerio, D., Overett, M. T., Malcolm, A., Blackburn, G. L. 1978. Nutritional support for cancer patients receiving abdominal and pelvic radiotherapy: A randomized prospective, clinical experiment of intravenous vs. oral feeding. *Surg. Forum* 29:145–48
64. Issell, B. F., Valdivieso, M., Zaren, H. A., Dudrick, S. J., Freirich, E. J., Copeland, E. W., Bodey, G. 1978. Protection against chemotherapy toxicity by I.V. hyperalimentation. *Cancer Treat. Rep.* 59:437–9
65. Heatley, R. V., Williams, R. H. P., Lewis, M. H. 1980. Preoperative intravenous feeding—a controlled trial. *Postgrad. Med. J.* In press
66. Popp, M. B., Fisher, R. I., Simon, R. M., Brennan, M. F. A prospective randomized study of adjuvant parenteral nutrition in the treatment of diffuse lymphoma: I. Effect on drug tolerance. *Cancer Treat. Rep.* In press
67. Serrou, B., Cupissol. D. 1980. Adjunct effect of parenteral intravenous nutrition (PIVN) depends on the tumor sensitivity to chemotherapy. *Am. Soc. Clin. Oncol.* C-157 (Abstr.)
68. Lanzotti, V., Copeland, E. M., Bhuchar, V., Wesley, M., Corriere, J., Dudrick, S. 1980. A randomized trial of total parenteral nutrition (TPN) with chemotherapy for non-oat cell cancer (NOLCLC) *Am. Soc. Clin. Oncol.* C-277 (Abstr.)
69. Moghissi, K., Hornshaw, J., Teasdale, P. R., Dawes, E. A. 1977. Parenteral nutrition in carcinoma of the oesophagus treated by surgery: nitrogen balance and clinical studies. *Br. J. Surg.* 64:125–28
70. Collins, J. P., Oxby, C. B., Hill, G. L. 1978. Intravenous amino acids and intravenous hyperalimentation as protein-sparing therapy after major surgery. *Lancet* 1:788–91
71. Lewins, R. 1831. Injection of saline solutions in extraordinary quantities into the veins in cases of malignant cholera. *Lancet* 2:243–44
72. Duma, R. J. 1980. Thomas Latta, what have we done? The hazards of intravenous therapy. *N. Engl. J. Med.* 294: 1178–80

Ann. Rev. Med. 1981. 32:245–59
Copyright © 1981 by Annual Reviews Inc. All rights reserved

MAGNESIUM DEFICIENCY AND EXCESS[1]

❖7396

R. K. Rude, M.D. and F. R. Singer, M.D.

Department of Medicine, University of Southern California School of Medicine,
and Orthopaedic Hospital/University of Southern California Bone and
Connective Tissue Research Laboratories, Los Angeles, California 90007

MAGNESIUM DEFICIENCY

Introduction

Since magnesium is a prominent element in the earth's crust and in sea water, it is not surprising that it is incorporated into the metabolic processes of all living things. In plants, magnesium is essential for the photobiosynthesis of ATP and the synthesis of plant carbohydrate. In animals, magnesium is essential for any enzymatic reaction requiring ATP, including oxidative phosphorylation. Therefore, DNA transcription, RNA aggregation, protein synthesis, and a host of plasma membrane functions depend on an optimal concentration of magnesium (1).

Magnesium is ubiquitous in food stuffs. This wide availability, along with efficient renal and gastrointestinal mechanisms controlling magnesium conservation, makes pure magnesium deficiency an uncommon entity. When it occurs it is usually associated with disease states whose symptoms may obscure the presence of magnesium deficiency. This is probably why magnesium deficiency was not recognized as a potential clinical problem until 1932 when Kruse et al (2) described experimental magnesium deficiency in the rat. This study was quickly followed by reports of clinical symptoms in patients with hypomagnesemia (3). Since that time experimental studies in animals and humans have clearly defined the physiological consequences (4, 5) of magnesium depletion.

Definition

In the last review of magnesium deficiency in this publication, Gitelman & Welt (6) stressed the difficulty in precisely defining a state of magnesium

---

[1] Suported in part by Grant CIA NIAM DD 1 k08AM00371-01.

0066-4219/81/0401-0245$01.00

deficiency. This problem still exists. Magnesium is the fourth most abundant cation in the body, but over 99% is either intracellular or in the skeleton. Thus the measurement of plasma magnesium concentration may not reflect the true total body magnesium content (7). Muscle, erythrocyte, and bone magnesium content also have been measured to assess their value as indices of magnesium deficiency. Muscle magnesium content decreases in magnesium depletion (8), but not rapidly (9, 10), and it does not correlate well with either plasma, erythrocyte, or bone magnesium content (10). Erythrocyte magnesium concentration appears to decrease only after prolonged magnesium depletion (4, 9–11). This suggests that erythrocyte magnesium content depends mainly on the extracellular magnesium concentration during hematopoiesis (11). Magnesium in bone, which accounts for about 60% of the total body magnesium, is either incorporated into the mineral lattice or remains in an elutable surface-limited mineral pool (12). This latter fraction accounts for about one third of the magnesium in bone (13). It is this fraction of bone magnesium that best correlates with the plasma magnesium concentration and may serve as a reservoir for the maintenance of a normal plasma magnesium concentration (12, 13). Since this pool of magnesium is difficult to assess, the plasma magnesium concentration remains the index by which most clinicians determine whether or not a state of magnesium deficiency exists.

The total plasma magnesium concentration in man, of which 65–75% is diffusible, is usually 1.5–2.0 meq/liter and appears to be tightly held within this range. A plasma concentration of less than 1.5 meq/liter presumably indicates some degree of magnesium depletion. Magnesium deficiency has also been documented in patients with levels above 1.5 meq/liter (7, 15).

In patients with borderline evidence of magnesium deficiency it may be useful to evaluate the amount of magnesium excreted in the urine following an infusion of magnesium (14). Normal subjects excrete at least 80% of an intravenous magnesium load in the urine within 24 hours, whereas patients with magnesium deficiency excrete less.

Disorders Causing Magnesium Deficiency

Magnesium deficiency is usually secondary to excessive losses of magnesium from the body. This loss most commonly occurs from the kidney or gastrointestinal tract. Nutritional magnesium deficiency is rarely if ever seen, but reducing dietary magnesium during acute and chronic illness will hasten the onset of symptomatic magnesium deficiency. Table 1 outlines the causes of magnesium deficiency.

GASTROINTESTINAL Magnesium appears to be absorbed along the entire small intestine and also to some extent in the large bowel (16). Early

Table 1 Causes of magnesium deficiency

A. Gastrointestinal disorders
  1. Malabsorption syndromes
  2. Extensive bowel resection
  3. Intestinal and biliary fistulas
  4. Chronic diarrhea
  5. Acute pancreatitis
  6. Protein-calorie malnutrition
  7. Primary hypomagnesemia

B. Renal loss
  1. Chronic parenteral fluid therapy
  2. Hypercalcemia
  3. Primary aldosteronism
  4. Drugs
    a. Diuretics
    b. Cisplatin
    c. Gentamicine
    d. Cardiac glycosides
  5. Osmotic diuresis
    a. Glucose
    b. Mannitol
    c. Urea

B. Renal loss (continued)
  6. Alcohol
  7. Starvation
  8. Renal diseases
    a. Chronic pylonephritis and glomerulonephritis
    b. Diuretic phase of ATN
    c. Postobstructive nephropathy
    d. Renal tubular acidosis
    e. Postrenal transplant
  9. Primary hypomagnesemia

C. Endocrine and metabolic
  1. Hungry bone syndrome
  2. Hypoparathyroidism
  3. Hyperthyroidism
  4. Hypothyroidism
  5. Phosphate deficiency
  6. Excessive lactation
  7. Protein caloric malnutrition/low protein diet

studies in rats suggested that the distal small bowel is the major site of magnesium absorption (17). These data correlate well with a high incidence of magnesium deficiency in disease of the distal small bowel in man. Resection or bypass of the ileum for various reasons, such as obesity, enteritis, or vascular infarction, frequently causes magnesium deficiency (18–21). Malabsorption syndrome secondary to nontropical sprue (21, 22) is associated with hypomagnesemia in 66% of cases (23). Radiation damage to the intestinal mucosa, Whipple's disease, and intestinal lymphectasia may also result in magnesium depletion (15, 21, 24). Steatorrhea may cause or potentiate magnesium malabsorption through formation of nonabsorbable magnesium-lipid salts (21).

Excessive loss of gastrointestinal fluids may create a significant negative magnesium balance (14). While gastric and biliary fluid contain relatively little magnesium (1–2 meq/liter), loss of large volumes from nasogastric suction or intestinal and biliary fistulae produces magnesium deficiency (7, 14, 24, 25). Secretions from the lower intestinal tract may be quite high in magnesium content (10–15 meq/liter) (7, 14). Prolonged diarrhea complicating ulcerative colitis, regional enteritis, laxative abuse, intestinal infections, and protein calorie malnutrition may cause profound magnesium deficiency (26–29).

A selective primary defect in the intestinal absorption of magnesium is a rare cause of magnesium deficiency (30, 31). This disorder becomes apparent during infancy and usually is diagnosed after the onset of hypocalcemia and seizures.

Acute hemorrhagic or edematous pancreatitis may produce hypomagnesemia (32). The pathogenesis of this is unclear but could be related to deposition of magnesium in the necrotic peripancreatic fat.

RENAL The kidney is the primary organ concerned with the fine regulation of magnesium metabolism. During magnesium deprivation, urinary magnesium excretion decreases to less than 1 meq/day (33). The renal handling of magnesium in man involves glomerular filtration of the diffusible fraction of plasma magnesium and reclamation by tubular reabsorption. The major site of magnesium reabsorption is the thick ascending limb of Henle (50–70%), while 15–20% of filtered magnesium is reabsorbed at the proximal tubule (34).

Magnesium reabsorption appears to be linked with calcium and sodium. Infusion of either of these cations increases renal magnesium excretion (35, 36). It is not surprising then that diseases, associated with hypercalcemia and hypercalciuria may result in hypomagnesemia. Examples include primary hyperparathyroidism, hypercalcemia of malignancy, and vitamin D intoxication (1, 26, 37). Intravenous saline infusion as well as volume expansion due to intravenous fluids or primary aldosteronism, also enhances renal magnesium excretion (1, 24, 36).

A number of drugs may produce renal wasting of magnesium. Diuretics, particularly the loop diuretics furosemide and ethacrynic acid, may cause significant magnesium diuresis (1, 38, 39). Cardiac glycosides also cause an increase in urinary magnesium excretion. This may potentiate the magnesium loss caused by diuretic therapy. Renal tubular damage from gentamicin (41) and cisplatin (42) may result in a reversible defect in renal magnesium conservation. Osmotic diuresis secondary to hyperglycemia, mannitol, or urea also increases urinary magnesium excretion (39, 43).

Alcoholism is frequently associated with magnesium deficiency and hypomagnesemia (1, 39). While the pathogenesis is usually multifactorial, involving dietary deprivation and diarrhea, a rising blood alcohol level increases urinary magnesium excretion (44). The mechanism is unknown.

During periods of starvation or fasting the urinary excretion of magnesium increases (45). Associated with this loss is the development of ketosis and metabolic acidosis. The acidosis may conceivably impair the ability of the renal tubule to reabsorb magnesium. Release of magnesium from bone may also contribute to the hypermagnesuria (45).

A number of diseases that cause renal tubular damage also cause renal magnesium wasting. Prominent among these are postobstructive nephropathy, the resolving phase of acute tubular necrosis, chronic glomerulonephritis and pyelonephritis, and post renal transplantation (1, 46). Renal tubular acidosis may also cause magnesium loss (1, 39). A primary defect in the renal reabsorption of magnesium has been reported (47, 48). Accompanying features of this disorder were nephrocalcinosis and hypokalemia.

ENDOCRINE AND METABOLIC Primary hyperparathyroidism may occasionally be associated with hypomagnesemia (37, 49, 50). As discussed above, this is probably secondary to the effect of calcium on the renal handling of magnesium. Profound hypomagnesemia with hypocalcemia and hypophosphatemia may be seen following surgical correction of hyperparathyroidism (51, 52). This is thought to be due to the "hungry bone syndrome," the healing phase of osteitis fibrosa during which increased amounts of calcium, magnesium, and phosphate are deposited in bone.

Hypomagnesemia is an infrequent finding in patients with hypoparathyroidism (53). The hypomagnesemia might result from renal magnesium wasting in the absence of parathyroid hormone, a hypothesis based on the observation that parathyroid extract acutely decreases urinary magnesium excretion in hypoparathyroid patients (54). Recent evidence indicating that endogenous parathyroid hormone does not alter renal handling of magnesium is at odds with this hypothesis (55).

Magnesium metabolism is altered in hyperthyroidism and hypothyroidism (56). Thyrotoxic patients have a decreased serum magnesium concentration along with increased urinary magnesium excretion. Intracellular magnesium concentration is normal. Conversely, hypothyroidism is associated with a mean increase in the serum magnesium concentration, decreased urinary magnesium excretion, and a low intracellular magnesium concentration (56). The mechanisms responsible for these changes are unknown.

Experimental phosphate deficiency has been shown to result in hypomagnesemia in the rat and man (57, 58). The mechanism is apparently related to increased urinary magnesium excretion. Hypophosphatemia is common in clinical medicine and may therefore play an important role in the magnesium deficiency syndrome in man.

An important component of protein-calorie malnutrition in children is the development of magnesium depletion. These children may be either hypomagnesemic or normomagnesemic with low urinary magnesium excretion. Muscle magnesium content is very low (29). The mechanism by which the magnesium depletion occurs is unknown but is probably related to dietary magnesium deprivation and attacks of gastroenteritis (29). Hypomagnesemic tetany has been described in one woman with excessive lactation of 2400 ml of milk per day (59).

Manifestations

BIOCHEMICAL Hypomagnesemia is the index by which a diagnosis of magnesium deficiency is made. Although about 30% of magnesium in the circulation is bound to albumin, corrections of the total plasma magnesium concentration in hypoalbuminemic patients are not generally made. There is no routine method for measuring ionized magnesium concentration.

It is widely recognized that moderate to severe magnesium deficiency may result in hypocalcemia in most species studied (1, 5, 15). An exception is the rat, in which hypercalcemia develops if the animal is on a normal calcium diet (1, 5). The pathogenesis of the hypocalcemia is of great interest. Impaired function of the parathyroid gland was implicated based on the observation that circulating immunoreactive parathyroid hormone (IPTH) was low in a number of magnesium-deficient hypocalcemic patients (15). This concept was confirmed by studying the levels of IPTH after rapid normalization of the serum magnesium concentration (60). An intravenous injection of magnesium produced a marked increase in IPTH concentration within minutes (60). An example of such a response in a magnesium-deficient hypocalcemic patient is shown in Figure 1. Similar responses were found in magnesium-deficient hypocalcemic patients in whom basal IPTH levels were normal or elevated (60). This suggests that impaired parathyroid hormone (PTH) secretion is a consistent abnormality in these patients. The immediate release of PTH after magnesium injection indicates that the biosynthesis of PTH is not abnormal.

End-organ resistance to PTH has also been demonstrated in hypocalcemic magnesium deficiency (5, 15). Impaired generation of cyclic AMP as determined by the urinary excretion of cyclic AMP in response to parathyroid extract has been shown in some, but not all, patients with severe magnesium deficiency (15). The response to parathyroid extract seems to depend on the severity of magnesium depletion. Skeletal resistance to PTH has also been proposed as a factor contributing to the hypocalcemia. The finding of normal or increased serum IPTH concentrations in hypocalcemic magnesium-deficient patients supports this contention (15, 60). A subnormal calcemic response to parathyroid extract has been reported in a number of such patients (61, 62). The concept of skeletal resistance to PTH action is further supported by in vitro studies demonstrating that PTH produced submaximal release of calcium from bones of magnesium-deficient rats (63).

The nature of the mechanisms by which magnesium deficiency impairs PTH secretion and action has been sought in the adenylate cyclase-cyclic AMP system. The stimulation of PTH secretion by hypocalcemia is mediated by plasma-membrane-bound adenylate cyclase (64). Recently it was demonstrated that the human parathyroid gland adenylate cyclase is mag-

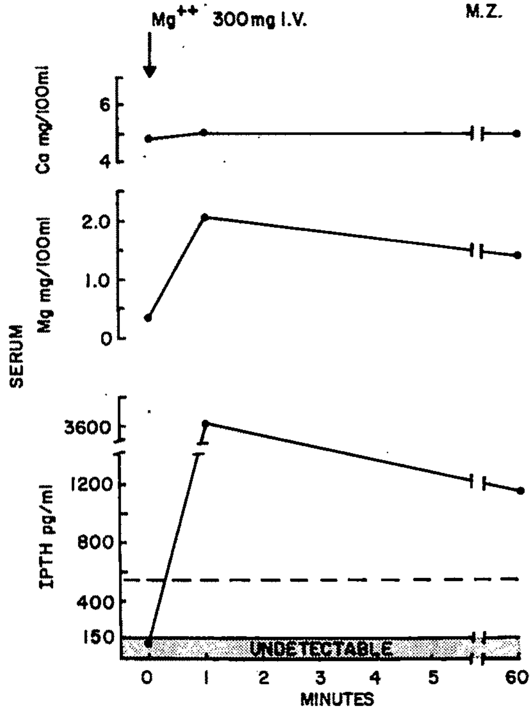


Figure 1 The effect of an intravenous injection of 300 mg elemental magnesium on the plasma concentration of calcium, magnesium, and IPTH in a hypocalcemic magnesium deficient patient. (Reprinted from Reference 15 with permission of publisher.)

nesium dependent (65). Evidence for depressed activity of renal adenylate cyclase has been provided by studies which demonstrated that urinary cyclic AMP excretion was reduced in response to parathyroid extract administration (15). Skeletal resistance to PTH could also be explained by reduced adenylate cyclase activity, since in isolated perfused bone from magnesium-deficient dogs there was impaired generation of cyclic AMP (66).

Finally, the hypocalcemia of magnesium deficiency is resistant to the action of vitamin D (67, 68). The nature of this resistance is not known but perhaps a defect in vitamin D metabolism and/or action is a contributing factor.

Hypokalemia is a frequent finding in magnesium deficiency (4, 47). It develops secondary to impaired ability of the kidney to reabsorb K^+ (4). Renal K^+ wasting and decreased intracellular K^+ (4, 9, 69) are thought to be related to an impaired ability of the Na-K pump to maintain normal intracellular concentrations of K^+ and Na^+. The impairment of ion transport is probably explained by the magnesium dependency of adenosine triphosphatase, an enzyme required for normal function of the Na-K pump (70).

NEUROMUSCULAR Patients with magnesium deficiency usually present with symptoms and signs involving the neuromuscular system (1, 4, 5, 19–21, 24, 25, 27, 28, 30, 31, 39, 47, 48). Positive Chvostek's and Trousseau's signs and frank tetany are common (24–28, 59). Since severe magnesium deficiency is frequently associated with hypocalcemia, it has been suggested that these manifestations of neuromuscular irritability may be secondary to the hypocalcemia. A report of hypomagnesemic normocalcemic patients with tetany documents that magnesium deficiency per se can cause this problem (28). Muscle tremors, fasciculations, wasting, and weakness may occur in chronic magnesium deficiency (24, 25, 27, 71). Although infrequently seen, athetoid and choreoiform movements are manifestations of magnesium deficiency (24, 25, 27, 71). Central nervous system symptoms such as nystagmus, ataxia, and vertigo as well as psychiatric abnormalities have been reported (24, 26, 27, 71)

CARDIOVASCULAR An important and serious clinical manifestation of magnesium deficiency is cardiac arrhythmia (72–76). Chronic magnesium deficiency results in changes in the electrocardiogram such as prolonged P-R and Q-T intervals and flat, broad T waves (74, 77). These changes are similar to those seen in hypokalemia, and lead one to speculate that these changes are secondary to the effect of magnesium depletion on potassium metabolism (76, 78) as discussed above. Arrhythmias such as premature ventricular beats, ventricular tachycardia, and ventricular fibrillation have been reported (72–76). The potential role of magnesium deficiency in triggering such arrhythmias must be kept in mind since standard antiarrhythmic drug therapy may be ineffective in controlling ventricular arrhythmias associated with magnesium deficiency (72, 73, 76). In some cases, only magnesium therapy will terminate the arrhythmia (72, 73).

An increased susceptibility to digitalis intoxication has been observed both clinically and experimentally in magnesium deficiency (79–81). It has been suggested that this is due to an enhanced effect of digitalis on intracellular potassium depletion (81). Both digitalis administration and magnesium deficiency reduce intracellular K^+ by inhibiting the Na-K pump (81).

GASTROINTESTINAL Dysphagia has recently been associated with magnesium deficiency (82, 83). The dysphagia is completely reversible with magnesium repletion. An increase in neuronal excitability and neuromuscular transmission has been suggested as the cause of the esophageal spasm (82).

HEMATOLOGIC Magnesium deficiency may result in an anemia characterized by a shortened erythrocyte survival, reticulocytosis, spherocytosis, microcytosis, and erythroid hyperplasia of the bone marrow (84). This has been demonstrated in experimental magnesium deficiency in man and animal, as recently summarized (84). The mechanism of the anemia is unclear but has been attributed to both increased destruction of erythrocytes in the spleen and to intravascular hemolysis.

Therapy

Symptomatic magnesium deficiency should be treated with magnesium salts. Patients with mild deficiencies may receive sufficient magnesium from their diet if they are able to eat. It should be recognized that the correction of hypocalcemia and hypokalemia along with resolution of symptoms of magnesium deficiency require therapy with magnesium. Treatment with vitamin D, calcium, or potassium will not correct the mineral imbalance (30, 31, 67, 68, 78). Prior to the administration of magnesium salts, renal function should be assessed. If the patient has a lowered glomerular filtration rate, the replacement dose of magnesium must be reduced as magnesium intoxication may result. Daily serum magnesium concentrations must be determined if renal function is impaired.

An effective regimen of magnesium replacement is 1.0 g of $MgSO_4 \cdot 7H_2O$ (8.13 meq Mg) every 4–6 hours for five days. If the intramuscular route is to be used, a 50% solution is available. If the patient requires intravenous infusions, the magnesium can be given in the infusion throughout the day. This is tolerated better than the painful, intramuscular magnesium injections. If necessary up to 100 meq/day for five days may be safely administered to patients with normal renal function. In our experience a dose of 32–48 meq magnesium per day will maintain a serum magnesium concentration at 2–2.5 mg/dl in most magnesium-deficient patients (15).

Magnesium equilibrates slowly with the intracellular compartment (11). In most hypocalcemic patients the serum calcium returns to normal four to five days after the onset of magnesium therapy (15). Therefore therapy should continue for at least a five-day period. If the patient after this time is still not eating and/or there is continuing magnesium loss, a daily maintainence dose of 8–10 meq magnesium should be instituted.

When magnesium losses are continuous, oral magnesium supplements in the form of sulfate, lactate, hydroxide, chloride, or glycerophosphate can be given. The amount of magnesium required will depend on the extent and route of loss. The symptoms of magnesium deficiency have been controlled by 10–160 meq magnesium daily (30, 31, 43).

MAGNESIUM EXCESS

Introduction

Symptomatic hypermagnesemia is an uncommon clinical problem although mild hypermagnesemia is not rare in a hospital population. Magnesium intoxication is most likely to occur when magnesium is administered as a therapeutic agent (1, 3, 85, 86).

Causes of Hypermagnesemia

EXCESSIVE INTAKE The oral ingestion of magnesium has rarely been reported to result in magnesium intoxication without coexisting renal failure. Rectal administration of magnesium may inexplicably cause hypermagnesemia in patients with normal renal function (85).

IMPAIRED EXCRETION The kidney is the major organ concerned with regulating magnesium metabolism. As glomerular filtration rate falls, urinary magnesium excretion also decreases (87). Hypermagnesemia is usually seen in patients with chronic renal failure when the glomerular filtration rate falls below 10–30 ml/min (3, 85, 87). Magnesium intoxication is most common in patients with chronic renal failure who have been receiving magnesium-containing antacids, enemas, or infusions (3, 30, 85, 87). Excessive dialysate magnesium also reportedly causes symptomatic hypermagnesemia (85).

Acute renal failure has also been associated with hypermagnesemia. The mean maximal serum magnesium values reported vary from 2.6–3.8 meq/liter. Azotemia, acidosis, rhabdomyolysis, and continued magnesium intake appear to be contributing factors (85).

PARENTERAL ADMINISTRATION Inappropriate administration of magnesium is infrequently encountered. Intraperitoneal magnesium given inadvertantly to two patients with a perforated viscus resulted in magnesium intoxication (85). Parenteral magnesium therapy is the treatment of choice in toxemia of pregnancy. Optimal blood levels of 4–7 meq/liter may cause symptoms of magnesium excess (1, 85). Overzealous therapy and/or inadequate monitoring of the plasma magnesium concentration could result in severe magnesium intoxication. Hypermagnesemia in the newborn has been reported following the administration of magnesium in the treatment of toxemia of pregnancy (25, 86).

OTHER CAUSES Mild to moderate degrees of hypermagnesemia occur in lithium ingestion, hypothyroidism, milk-alkali syndrome, viral hepatitis, and Addison's disease (for review, see 85).

Manifestations

BIOCHEMICAL The plasma magnesium concentration usually exceeds 4 meq/liter before any signs or symptoms of magnesium excess appear (1, 25, 85). Mild to moderate hypermagnesemia may remain unnoticed unless a high index of suspicion exists and a plasma magnesium concentration is determined. At magnesium concentrations greater than 4 meq/liter the manifestations of hypermagnesemia as described below may develop.

Parenteral administration of magnesium lowers plasma calcium concentration in both normal and hypoparathyroid man (53, 85, 86). Suppressed parathyroid hormone secretion and/or impairment of the peripheral effects of parathyroid hormone may be the major factors responsible for the hypocalcemia in parathyroid intact subjects (86, 88). The reduction of plasma calcium in hypoparathyroid patients is unexplained. Excess magnesium has been reported to impair clotting and may play a role in uremic clotting disorders (85).

NEUROMUSCULAR Hypermagnesemia may result in symptoms referable to the neuromuscular system (3, 25, 26, 85, 89). Excessive magnesium decreases the impulse transmission across the neuromuscular junction. A decrease in postsynaptic membrane responsiveness and an increase in the threshold of axonal excitation also impair neuromuscular function (85, 89). Clinically, one of the earliest effects of magnesium intoxication is a decrease or disappearance of the deep tendon reflexes which may occur at plasma magnesium levels of 4 meq/liter (1, 25, 85). Somnolence is seen at levels of 4–7 meq/liter and paralysis of voluntary muscles at 10 meq/liter or greater. The latter may impair respiratory function and during severe magnesium intoxication apnea may occur (1, 25). As a general guideline, when the deep tendon reflexes are absent, respirations should be closely monitored

CARDIOVASCULAR Various electrocardiographic changes are associated with elevations in the plasma magnesium concentration. An increase in the P-R interval and intraventricular conduction defects with increased QRS duration and QT interval occur at plasma magnesium concentrations of 5–10 meq/liter. Complete heart block may occur at levels greater than 15 meq/liter. High concentrations of magnesium inhibit cardiac contractions and decrease membrane excitability. Cardiac arrest in asystole may occur at plasma magnesium concentrations of 15 meq/liter (85, 89). Moderate elevation in the plasma magnesium concentration (4–5 meq/liter) may result in bradycardia and mild supine and orthostatic hypotension. Proposed mechanisms have recently been reviewed (85).

OTHER Early nonspecific manifestations of magnesium intoxication include nausea, vomiting, and cutaneous flushing. These symptoms may appear at plasma magnesium concentrations of 3–9 meq/liter.

Therapy

The plasma magnesium concentration should be monitored in those patients at greatest risk for the development of hypermagnesemia; that is, patients with renal failure and/or patients receiving magnesium-containing medications. Patients with mild to moderate hypermagnesemia who are receiving magnesium should have the medication discontinued. Renal excretion of magnesium will decrease the plasma magnesium concentration and resolve the symptoms of magnesium excess. Patients with higher elevations of the plasma magnesium and severe symptoms of magnesium intoxication may require more aggressive therapy. Calcium will transiently reverse the symptoms of magnesium excess. The usual dose is a 5–10 minute infusion of 100–200 mg of elemental calcium as either the gluconate or chloride salt (90). Peritoneal dialysis or hemodialysis will rapidly and effectively lower the plasma magnesium concentration. This mode of therapy is especially beneficial if renal function is impaired (85, 90).

ACKNOWLEDGMENTS

We wish to thank Mrs. Allana Stephens for her secretarial assistance.

Literature Cited

1. Wacker, W. E. C., Parisi, A. F. 1968. Magnesium metabolism. *N. Engl. J. Med.* 278:658–63, 712–17, 772–76
2. Kruse, H. D., Orent, E. R., McCollum, E. V. 1932. Studies on magnesium deficiency in animals I. Symptomatology resulting from magnesium deprivation. *J. Biol. Chem.* 96:519–39
3. Hirschfelder, A. D., 1934. Clinical manifestations of high and low plasma magnesium. *J. Am. Med. Assoc.* 102:1138–41
4. Shils, M. E. 1969. Experimental human magnesium depletion. *Medicine* 48:61–82
5. Shils, M. E. 1976. Magnesium deficiency and calcium and parathyroid hormone interrelations. In *Trace Elements in Human Health and Disease,* ed. A. S. Prasad, 27:23–46. New York, San Francisco, London: Academic. 525 pp.
6. Gitelman, H. J., Welt, L. G. 1969. Magnesium deficiency. *Ann. Rev. Med.* 20:233–42
7. Barnes, B. A. 1969. Magnesium conservation: a study of surgical patients. *Ann. NY Acad. Sci.* 162:786–801
8. Booth, C. C., Hanna, S., Babouris, N., MacIntyre, I. 1963. Incidence of hypomagnesaemia in intestinal malabsorption. *Br. Med. J.* 1:141–44
9. Cotlove, E., Holliday, M. A., Schwartz, R., Wallace, W. M. 1951. Effects of electrolyte depletion and acid-base disturbance on muscle cations. *Am. J. Physiol.* 167:665–74
10. Elin, R. J., Armstrong, W. D., Singer, L. 1971. Body fluid electrolyte composi-

tion of chronically magnesium-deficient and control rats. *Am. J. Physiol.* 220:543–48

11. Watson, W. S., Hilditch, T. E., Horton, P. W. Davies, D. L., Lindsay, R. 1979. Magnesium metabolism in blood and the whole body in man using ^{28}Magnesium. *Metabolism* 28:90–95
12. Alfrey, A. C., Miller, N. L., 1973. Bone magnesium pools in uremia. *J. Clin. Invest.* 52:3019–27
13. Alfrey, A. C., Miller N. L., Butkus, D. 1974. Evaluation of body magnesium stores. *J. Lab. Clin. Med.* 84:153–62
14. Thorén, L. 1963. Magnesium deficiency in gastrointestinal fluid loss. *Acta Chir. Scand. Suppl.* 306:5–60
15. Rude, R. K., Oldham, S. B., Singer, F. R. 1976. Functional hypoparathyroidism and parathyroid hormone end-organ resistance in human magnesium deficiency. *Clin. Endocrinol.* 5: 209–24
16. Brannan, P. G., Vergne-Marini, P., Pak, C. Y. C., Hull, A. R., Fordtran J. S. 1976. Magnesium absorption in the human small intestine: results in normal subjects, patients with chronic renal disease, and patients with absorptive hypercalciuria. *J. Clin. Invest.* 57:1412–18
17. MacIntyre, I., Robinson, C. J. 1969. Magnesium and the gut: experimental and clinical observations. *Ann. NY Acad. Sci.* 162:865–63
18. Swenson, S. A. Jr., Lewis, J. W., Sebby, K. R. 1974. Magnesium metabolism in man with special reference to jejunoileal bypass for obesity. *Am. J. Surg.* 127: 250–55
19. Opie, L. H., Hunt, B. G., Finlay, J. M. 1964. Massive small bowel resection with malabsorption and negative magnesium balance. *Gastroenterology* 47:415–20
20. Heaton, F. W., Fourman, P. 1965. Magnesium deficiency and hypocalcaemia in intestinal malabsorption. *Lancet* 2: 50–52
21. Booth, C. C., Hanna, S., Babouris, N., MacIntyre, I. 1963. Incidence of hypomagnesaemia in intestinal malabsorption. *Br. Med. J.* 2:141–44
22. Goldman, A. S., Van Fossan, D. D., Baird, E. E. 1962. Magnesium deficiency in celiac disease. *Pediatrics* 29:948–52
23. Arnaud, S. B., Newcomer, A. D., Go, V. L. W. 1976. Hypomagnesemia and vitamin D deficiency in patients with non-tropical sprue. *Clin. Res* 24:564A
24. Vallee, B. L., Wacker, W. E. C., Ulmer, D. D. 1960. The magnesium-deficiency tetany syndrome in man. *N. Engl. J. Med.* 262:155–61
25. Fishman, R. A. 1965. Neurological aspects of magnesium metabolism. *Arch. Neurol.* 12:562–69
26. Hanna, S. 1961. Plasma magnesium in health and disease. *J. Clin. Pathol.* 14:410–14
27. Hanna, S., Harrison, M., MacIntyre, I., Fraser, R. 1960. The syndrome of magnesium deficiency in man. *Lancet* 2:172–76
28. Wacker, W. E. C., Moore, F. D., Ulmer, D. D., Vallee, B. L., 1962. Normocalcemic magnesium deficiency tetany. *J. Am. Med. Assoc.* 180:161–63
29. Caddell, J. L. 1969. Magnesium deficiency in protein-calorie malnutrition: a follow-up study. *Ann. NY Acad. Sci.* 162:874–90
30. Stromme, J. H., Nesbakken, R., Normann, T. Skjorten, F., Skyberg, D., Johannessen, B. 1969. Familial hypomagnesemia. *Acta Paediatr. Scand.* 58: 433–44
31. Nordio, S., Donath, A., Macagno, F., Gatte, R. 1971. Chronic hypomagnesemia with magnesium-dependent hypocalcemia. *Acta Paediatr. Scand.* 60:441–48
32. Edmondson, H. A., Berne, C. J., Homann, R. E. Jr., Wertman, M. 1952. Calcium, potassium, magnesium and amylase disturbances in acute pancreatitis. *Am. J. Med.* 12:34–42
33. Heaton, F. W. 1969. The kidney and magnesium homeostasis. *Ann. NY Acad. Sci.* 169:775–85
34. Carney, S. L., Wong, N. L. M., Quamme, G. A., Dirks, J. H. 1980. Effect of magnesium deficiency on renal magnesium and calcium transport in the rat. *J. Clin Invest.* 65:180–88
35. Coburn, J. W., Massry, S. G., Kleeman, C. R. 1970. The effect of calcium infusion on renal handling of magnesium with normal and reduced glomerular filtration rate. *Nephron* 7:131–43
36. Massry, S. G., Coburn, J. W., Chapman, L. W., Kleeman, R. 1967. Effect of NaCl infusion on urinary Ca and Mg during reduction in their filtered loads. *Am. J. Physiol.* 213:1218–24
37. Walser, M. 1962. Separate effects of hyperparathyroidism, hypercalcemia of malignancy, renal failure, and acidosis on state of calcium, phosphate, and other ions in plasma. *J. Clin. Invest.* 41:1454–71

38. Duarte, C. G. 1968. Effects of ethacrynic acid and furosemide on urinary calcium, phosphate, and magnesium. *Metabolism* 17:867–76
39. Martin, H. E. 1969. Clinical magnesium deficiency. *Ann. NY Acad. Sci.* 169:891–900
40. Kupfer, S., Kosovsky, J. D. 1965. Effects of cardiac glycosides on renal tubular transport of calcium, magnesium, inorganic phosphate, and glucose in the dog. *J. Clin. Invest.* 44:1132–42
41. Bar, S. R., Wilson, H. E., Mazzaferri, E. L. 1975. Hypomagnesemic hypocalcemia secondary to renal magnesium wasting. *Ann. Intern. Med.* 82:646–49
42. Schilsky, R. L., Anderson, T. 1979. Hypomagnesemia and renal magnesium wasting in patients receiving cisplatin. *Ann. Intern. Med.* 90:929–31
43. Massry, S. G. 1977. Pharmacology of magnesium. *Ann. Rev. Pharmacol.* 17:67–82
44. Mendelson, J. H., Ogata, M., Mello N. K. 1969. Effects of alcohol ingestion and withdrawal on magnesium states of alcoholics: clinical and experimental findings. *Ann. NY Acad. Sci.* 169: 918–33
45. Drenick, E. J., Hunt, I. F., Swendseid, M. E. 1969. Magnesium depletion during prolonged fasting of obese males. *J. Clin. Endocrinol. Metab.* 29:1341–48
46. Davis, B. B., Preuss, H. G., Murdaugh, H. V. Jr. 1975. Hypomagnesemia following the diureses of post-renal obstruction and renal transplant. *Nephron* 14:275–80
47. Gitelman, H. J., Graham, J. B., Welt, L. G. 1966. A new familial disorder characterized by hypokalemia and hypomagnesemia. *Trans. Assoc. Am. Physicians* 79:221–35
48. Runeberg, L., Collan, Y., Jokinen, E. J., Lahdevirta, J., Aro, A. 1975. Hypomagnesemia due to renal disease of unknown etiology. *Am. J. Med.* 59:873–81
49. Sutton, R. A. L. 1970. Plasma magnesium concentration in primary hyperparathyroidism. *Br. Med. J.* 1:529–33
50. King, R. G., Stanbury, S. W. 1970. Magnesium metabolism in primary hyperparathyroidism. *Clin. Sci.* 39: 281–303
51. Heaton, F. W., Pyrah, L. N. 1963, Magnesium metabolism in patients with parathyroid disorders. *Clin. Sci.* 25: 475–85
52. Jones, C. T. A., Sellwood, R. A., Evanson, J. M. 1973. Symptomatic hypomagnesaemia after parathyroidectomy. *Br. Med. J.* 3:391–92
53. Jones, J. H., Fourman, P. 1966. Effects of infusions of magnesium and of calcium in parathyroid insufficiency. *Clin. Sci.* 30:139–50
54. Bethune, J. E., Turpin, R. A., Inoue, H., 1968. Effect of parathyroid hormone extract on divalent ion excretion in man. *J. Clin. Endocrinol. Metab.* 28:673–78
55. Rude, R. K., Bethune, J. E., Singer, F. R. 1980. Renal tubular maximum for magnesium in normal, hypoparathyroid, and hyperparathyroid man. *J. Clin. Endocrinol. Metab.* 51: In press
56. Jones, J. E., Desper, P. C., Shane, S. R., Flink, E. B. 1966. Magnesium metabolism in hyperthyroidism and hypothyroidism. *J. Clin. Invest.* 45:891–900
57. Dominquez, J. H., Gray, R. W., Lemann, J. Jr. 1976. Dietary phosphate deprivation in women and men: effects on mineral and acid balances, parathyroid hormone and the metabolism of 25-OH-vitamin D. *J. Clin. Endocrinol. Metab.* 43:1056–68
58. Kreusser, W., Kurakawa, K., Aznar, E., Sachtjen, E., Massry, S. G. 1978. Effect of PO_4 depletion on magnesium metabolism in rats. *J. Clin. Invest.* 61:573–81
59. Greenwald, J. H., Dubin, A., Cardon, L. 1963. Hypomagnesemic tetany due to excessive lactation. *Am. J. Med.* 35:854–59
60. Rude, R. K., Oldham, S. B., Sharp, C. F., Singer, F. R. 1978. Parathyroid hormone secretion in magnesium deficiency. *J. Clin. Endocrinol. Metab.* 47:800–6
61. Estep, H., Shaw, W. A., Watlington, C., Hobe, R., Holland, W., Tucker, St. G. 1969. Hypocalcemia due to hypomagnesemia and reversible parathyroid hormone unresponsiveness. *J. Clin. Endocrinol. Metab.* 29:842–48
62. Woodard, J. C., Wevster, P. D., Carr, A. A. 1972. Primary hypocalcemia, diarrhea and insensitivity to parathyroid hormone. *Digest Dis.* 17:612–18
63. MacManus, J., Heaton, F. W., Lucas, P. W. 1971. A decreased response to parathyroid hormone in magnesium deficiency *J. Endocrinol.* 49:253–58
64. Abe, M., Sherwood, L. M. 1972. Regulation of parathyroid hormone secretion by adenyl cyclase. *Biochem. Biophys. Res. Commun.* 48:396–401
65. Rodriquez, H. J., Morrison, A., Slatopolsky, E., Klahr, S. 1978. Adenylate

cyclase of human parathyroid gland. *J. Clin. Endocrinol. Metab.* 47:319–25

66. Freitag, J. J., Martin, K. J., Conrades, M. B., Bellorin-Front, E., Teitelbaum, S., Klahr, S. Slatopolsky, E. 1979. Evidence for skeletal resistance to parathyroid hormone in magnesium deficiency. *J. Clin. Invest.* 64:1238–44
67. Lifshitz, F., Harrison, H. C., Harrison, H. E. 1967. Response to Vitamin D of magnesium deficient rats. *Proc. Soc. Exp. Biol. Med.* 125:472–76
68. Medalle, R., Waterhouse, C., Hahn, T. J. 1976. Vitamin D resistance in magnesium deficiency. *Am. J. Clin. Nutr.* 29:854–58
69. Whang, R., Morosi, H. J., Rodgers, D., Reyes, R. 1967. The influence of sustained magnesium deficiency on muscle potassium repletion. *J. Lab. Clin. Med.* 70:895–902
70. Seller, R. H., Cangiano, J., Kim, K. E., Mendelssohn, S., Brest, A. N., Swartz, C. 1970. Digitalis toxicity and hypomagnesemia. *Am. Heart J.* 79:57–68
71. Flink, E. B. 1956. Magnesium deficiency syndrome in man. *J. Am. Med. Assoc.* 102:1138–41
72. Chadda, K. D., Lichestein E., Gupta, P. 1973. Hypomagnesemia and refractory cardiac arrhythmia in a nondigitalized patient. *Am. J. Cardiol.* 31:98–100
73. Iseri, L. T., Freed, J., Bures, A. R. 1975. Magnesium deficiency and cardiac disorders. *Am. J. Med.* 58:837–46
74. Burch, G. E., Giles, T. D. 1977. The importance of magnesium deficiency in cardiovascular disease. *Am. Heart J.* 94:649–57
75. McMullen, J. K. 1977. Asystole and hypomagnesaemia during recovery from diabetes ketoacidosis. *Br. Med. J.* 1:690
76. Fisher, J., Abrams, J. 1977. Life-threatening ventricular tachyarrhythmias in delirium tremens. *Arch. Intern. Med.* 137:1238–41
77. Seelig, M. S. 1969. Electrographic patterns of magnesium depletion appearing in alcoholic heart disease. *Ann. NY Acad. Sci.* 162:906–17
78. Dyckner, T., Wester, P. O. 1979. Ventricular extrasystoles and intracellular electrolytes before and after potassium and magnesium infusions in patients on diuretic treatment. *Am. Heart J.* 97:12–17
79. Beller, G. A., Hood, W. B., Smith, T. W., Abelmann, W. H., Wacker, W. E. C. 1974. Correlation of serum magnesium levels and cardiac digitalis intoxication. *Am. J. Cardiol.* 33:225–29
80. Seller, R. H., Ramirez, O., Brest, A. N., Moyer, J. H. 1966. Serum and erythrocytic magnesium levels in congestive heart failure. *Am. J. Cardiol.* 17:786–91
81. Seller, R. H., Cangiano, J., Kim, K. E., Mendelssohn, S., Brest, A. N., Swartz, C. 1970. Digitalis toxicity and hypomagnesemia. *Am. Heart J.* 79:57–68
82. Hamed, I. A., Lindeman, R. D. 1978. Dysphagia and vertical nystagmus in magnesium deficiency. *Ann. Intern. Med.* 89:222–23
83. Flink, E. B. 1978. Dysphagia in magnesium deficiency. *Ann. Intern. Med.* 89:282
84. Elin, R. J., Alling, D. W. 1978. Survival of normal and magnesium-deficient erythrocytes in rats: effect of magnesium-deficient diet vs. splenectomy. *J. Lab. Clin. Med.* 91:666–72
85. Mordes, J. P., Wacker, W. E. 1978. Excess magnesium. *Pharmacol. Rev.* 29:273–300
86. Donovan, E. F., Tsang, R. C., Steichen, J. J., Strub, R. J., Chen, I. W. 1980. Neonatal hypermagnesemia: effect on parathyroid hormone and calcium homeostasis. *J. Pediatr.* 96:305–10
87. Popovtzer, M. M., Massry, S. G., Coburn, J. W., Kleeman, C. R. 1969. The interrelationship between sodium, calcium, and magnesium excretion in advanced renal failure. *J. Lab. Clin. Med.* 73:763–71
88. Slatopolsky, E., Mercado, A., Morrison, A., Yates, J., Klahr, S. 1976. Inhibitory effects of hypermagnesemia on the renal action of parathyroid hormone. *J. Clin. Invest.* 58:1273–79
89. Engbaek, L. 1952. The pharmacological actions of magnesium ions with particular reference to the neuromuscular and the cardiovascular system. *Pharmacol. Rev.* 4:396–414
90. Massry, S. G. 1980. Managing disorders of serum magnesium concentration. *Drug Ther. (Hosp.)* Feb:43–52

Ann. Rev. Med. 1981. 32:261–71

AMYLOIDOSIS[1]

◆7397

Peter D. Gorevic, M.D.

Division of Allergy, Rheumatology, and Clinical Immunology, State
University of New York, Stony Brook, New York 11794

Edward C. Franklin, M.D.

Irvington House Institute, Department of Medicine,
New York University, School of Medicine, 550 First Avenue New York,
New York 10016

Introduction

The term amyloidosis encompasses a heterogeneous group of disorders, some systemic and some peculiar to particular organ systems or cell types, some clearly pathologic and some concomitants of normal processes such as aging. Common denominators for these diverse conditions are structural peculiarities of certain proteins that render them amyloidogenic, as well as defective processing at the tissue level that results in the polymerization and proteolytic cleavage of these proteins. We outline here some features of amyloidogenesis that have become apparent from recent studies. The reader is referred to several reviews (1–3) and three international symposia (4–6) for further information.

Structural Components of Amyloid Fibrils

Tissue amyloid has traditionally been defined by its distinctive tinctorial and ultrastructural properties (Table 1). Electron microscopic analysis of amyloid deposits from a variety of sources show characteristic fibrils; in extracted amyloid, a minor component with a pentagonal configuration (P-component) has also been identified (Table 2).

---

[1] The US Government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this paper.

Table 1 Criteria for the definition of amyloid

1. Homogeneous hyaline extracellular material on routine sections.
2. Crystal violet metachromasia and positive congo red staining.
3. Apple green positive birefringence after staining with alkaline congo red.
4. Fibrillar ultrastructure; ?P-component (AP) in extracted material.
5. Forms opalescent suspensions in low ionic strength solutions, but precipitates in NaCl >0.0075–0.01 M.
6. Composed of low molecular weight subunit proteins, dissociable in dilute acid or alkali, urea formate, guanidine, or sodium dodecyl sulfate (SDS).

Early methods for extraction of crude amyloid preparations utilized sucrose gradients (7) or "top-layer" preparations after homogenization and differential centrifugation (8). In 1968, a method was described for the isolation of amyloid fibrils in high yield, based on their property of forming opalescent colloidal suspensions following homogenization in low ionic strength solutions (9). Studies of these preparations by x-ray diffraction (10) and infrared spectroscopy (11) revealed a predominantly beta-pleated sheet configuration of fibrillar proteins that is responsible for characteristic congo red binding and birefringence. Crystal violet metachromasia, on the other hand, appears to be due to mucopolysaccharides intimately associated with, but not an integral component of, the amyloid fibril (12). Subsequent biochemical studies revealed all fibrils, regardless of source, to be composed of low molecular weight subunit proteins that differ depending on the type of amyloid and may, in some instances, be derived from serum precursors (2). A nomenclature has been proposed to classify amyloid proteins derived from systemic, local, and heredofamilial variants of amyloidosis (6) (Table

Table 2 Ultrastructural components of tissue amyloid

	Fibril	P-component (AP)
Contribution	~95%	~5%
Appearance	100–150 Å diameter variable length	laterally stacked rods of variable length; ring of 5 globular subunits in cross section
Carbohydrate content	minimal	glycoprotein (~11%)
Solubility	water	saline
Immunogenicity	poor	good
Physical properties	tends to polymerize under physiologic conditions	calcium-dependent binding to polyanions

3). The P-component constitutes 5–10% of the dry weight of all amyloid substances in man and in all experimental animals. Ultrastructurally, it looks like a doughnut consisting of five subunits in cross-section, which, when stacked one on top of the other, assume a rod-like appearance (14). P-component corresponds to a serum alpha globulin with a molecular weight (MW) of about 220,000, which consists of ten subunits, each with a MW of approximately 23,000 (15–17). It has striking structural homologies to C-reactive protein (18), but unlike C-reactive protein it does not behave as an acute phase reactant (19). The tendency of P-component to bind to Sepharose® beads in the presence of calcium and to dissociate in EDTA is the basis of a procedure for its isolation from serum (20).

Light Chain-Related Amyloid

The major protein component associated with primary and myeloma-associated amyloid (AL) was shown to be derived from immunoglobulin light chains by three types of experiments carried out by Glenner and collaborators in the early 1970s (21). Amino acid sequence (22) and peptide map studies (23) showed that the AL amyloid subunit had striking structural homologies to immunoglobulin light chains, most commonly those of

Table 3 Clinicopathologic and biochemical nomenclatures for amyloidosis

Clinicopathologic	Fibril subunit protein (6)
Primary (idiopathic) myeloma-associated macroglobulinemia pulmonary (nodular)	AL (light chain related)
Secondary chronic infections chronic inflammatory diseases Familial Mediterranean Fever (FMF)	AA (nonimmunoglobulin)
Heredofamilial (excluding FMF)	AF (prealbumin in some neuropathic forms)
Endocrine medullary thyroid carcinoma diabetes mellitus	AEt (thyrocalcitonin) other prohormones (i.e. insulin and glucagon
Senile	
Cardiac	AS_{c1} AS_{c2} (isolated atrial and generalized variants)
Brain	AS_B
Cutaneous	AD

the λ type. Subsequently it was shown that trypsin, pepsin (24, 25), and lysosomal enzymes (26) can convert certain immunoglobulin light chains, especially those of the λ class, into fibrils with the appearance on electron microscopy of amyloid. Thirdly, immunological cross-reactivity between amyloid subunits and immunoglobulin light chains was demonstrable using antisera to the amyloid substance (27).

AA Protein Amyloid

The major subunit of the secondary type of amyloid, and at least one familial form, that associated with Familial Mediterranean Fever, is the AA protein. The AA protein, which has been fully sequenced in man (28) and several lower species, generally consists of 76 residues and has a molecular weight of 8500. It has no structural homology with any known protein. The amino acid sequence of amyloid deposits obtained from different species shows striking conservation (29). It is of interest that many AA proteins show heterogeneity at both the amino and carboxy terminal ends (28, 30). AA proteins obtained from patients with a variety of diseases have been found to be similar, but not identical. Most of the variability is located in the carboxy terminal region.

The origin of the AA protein remained somewhat of a mystery until the discovery of antisera that recognized the presence in serum of a closely related molecule known as the SAA (serum amyloid A related) protein (31). The SAA protein has an electrophoretic mobility of an α1-globulin and can be dissociated to a 12–14,000 dalton subunit (32), designated SAAL. SAAL exists in serum as a large complex with a MW variously estimated as 100–200,000 (31, 33). It appears to circulate in close association with other serum proteins, particularly certain high density lipoproteins (34) and perhaps also serum albumin (35). Since the amino acid sequences of AA and SAA are identical for the first 76 residues (36), it seems likely that SAA gives rise to the AA protein by proteolytic cleavage of the carboxy terminal fragment.

Other Protein Subunits

The other types of subunits are more difficult to purify; consequently, their nature has not been defined as precisely. Prealbumin has been identified in the Portuguese Familial form (37) and a precursor of thyrocalcitonin in medullary carcinoma of the thyroid (38). The amyloid deposits in the pancreas that occur with aging and diabetes have not been well characterized, but may be related to insulin and glucagon since these β-pleated sheet molecules can be converted to fibrils by proteolysis (39). The senile type of amyloid is difficult to study because of the small amounts present. The subunits in the senile deposits in the heart range in molecular weight from

5–12,000 and differ from the other well-characterized types of amyloid proteins (40, 41).

Amyloidogenic Proteins

Increased production of SAA or light chains is a prerequisite for, but does not necessarily result, in amyloid deposition (42, 43). Therefore, especially among L chains, there may be some whose variable region structure makes them more susceptible to the processing that leads to amyloidosis. Amyloid subunits can consist of fragments smaller than the variable region (22), intact Bence-Jones proteins (44, 45), or mixtures of fragments, always including the variable region (V_L) but with differing lengths of the constant region (45). Clinically, amyloidosis shows a strong association with λ chain proteins (22, 46, 47), and one subclass designated as Vλ VI is almost always associated with amyloid deposits (2).

The polymorphisms of SAA and AA proteins are much more subtle in different individuals. Although animal studies (48) warrant the speculation that some SAA molecules are more likely to be processed as fibrils than others, no sequence data in man has yet been obtained to corroborate this possibility, and it seems likely that differences in processing may be more important (see below).

Proteolysis, Polymerization, and Persistence

The proteases responsible for the conversion of SAA to AA have recently been partially characterized. In vitro, SAA degradation occurs in two steps, the first of which involves the formation of an intermediate with molecular weight and antigenic properties of AA protein. Degradation of the AA-like intermediate to smaller peptides is a separate event, and a defect in this step has been postulated in some instances to facilitate the accumulation of AA in tissue (49). The proteases responsible for both events are intimately associated with the cell surface of peripheral mononuclear cells and resemble elastases, based on the effect of specific inhibitors (50). Fibril formation may be due to polymerization (51) or binding to other serum or tissue ground substance proteins (34, 35). Though an amphipathic amino terminal helical segment (52) may be responsible in part for the association of SAA with high density lipoproteins (34), lipid is not a major constituent of fibril preparations (53). Proteolysis (24, 25) or thermal denaturation (54) of some light chains in vitro leads to the generation of the beta-pleated sheet fibrils. The resistance of the fibrils to proteolysis (55) and phagocytosis (56) and their poor immunogenicity (57) may explain the lack of tissue reaction around amyloid deposits as well as their persistence once formed.

Cellular Origin

AL proteins derive from cells of B cell lineage (plasma cells, lymphocytes) (46, 60–62). In experimental murine amyloidosis of the AA type, reticuloendothelial hyperplasia is an early feature. By contrast, in the later stages of the disease, tissue reaction around established deposits may be minimal (63). Replacement is partly responsible for T cell depletion (lymphopenia) and depressed cell-mediated immunity, noted in humans (64) and experimental animals (65). Local synthesis of fibrils by contiguous macrophages was suggested by Smetana (66) and supported by the extensive histochemical studies of Teilum (67). Short-term culture experiments using medullary thyroid carcinoma showed ongoing synthesis of amyloid in vitro (68). More recent work demonstrates the ability of splenic explants (69, 70) or monolayers (71) from casein-stimulated mice to produce amyloid fibrils, and possibly AA protein, in vitro. Of interest in this regard is the phenomenon of "transfer amyloidosis," in which it has been possible to considerably shorten the induction phase of casein amyloidosis by adoptive transfer of splenic macrophage (72). Endotoxin or casein acutely augment SAA production in animals (73) and etiocholanolone has the same effect in human volunteers (74). SAA release depends on the production of a macrophage-derived inducer and can be abrogated by inhibitors of protein synthesis (75, 76). Direct assay of organ homogenates following injection of endotoxin, coupled with immunohistological studies, implicated the hepatocyte as the major source of SAA (77, 79). This conclusion was supported by the observation that ethionine, a selective inhibitor of liver protein synthesis, was able to inhibit both SAA production and the appearance of tissue amyloid (78).

Diagnosis

Biopsy remains the mainstay of diagnosis. Anti-AA antiserum has proven particularly useful for immunohistological studies. Though some AL antisera only recognize their homologous protein in tissue section, many AL deposits can be recognized with appropriate anti L-chain sera (80). AEt and AS (Table 3) proteins are unique and, in the latter case, perhaps organ-specific (81). The recent observation that paraffin-embedded AA amyloid loses specific congo red affinity following treatment with potassium permanganate, whereas AL and localized (AEt, AS, etc) amyloids do not (82), permits the examination of fixed material for retrospective studies. Ultimately, antisera will prove of use in diagnosing the different types of amyloidosis in tissue section.

Therapy

It is important to recall that clinical manifestations result from organ replacement and not tissue reaction, making supportive care an important facet of treatment. Several reports have shown prolonged survival in patients with renal amyloid maintained by dialysis or transplantation (83). Other reports have noted regression of secondary amyloidosis following adequate treatment of an underlying disease (13, 84).

Colchicine and some immunosuppressives may inhibit precursor protein synthesis. The former is effective prophylactically in Familial Mediterranean Fever (85), but has only been sporadically employed for other types or established amyloid. It is important to note that SAA is a normal serum protein, not uniformly associated with tissue amyloid, that may in fact serve an immunoregulatory function (86, 87). Alternatively, agents such as colchicine or penicillamine may inhibit cells responsible for proteolysis or somehow enhance the tissue response that degrades and removes fibrils once formed.

Most commonly, however, the patient with systemic amyloid presents at an advanced stage of the disease. Use of immunosuppressive therapy at this point has yielded generally disappointing results (88), though newer combination regimens continue to be tested. The search for an agent capable of lysing established tissue deposits has recently focused on dimethylsulfoxide (DMSO), a drug capable of dissociating in vitro noncovalently bonded subunit proteins and of retarding experimental amyloidosis (89). Case reports have shown clinical reponse in some patients with advanced disease (90), but the overall efficacy and safety of this drug remain to be established.

Conclusions

Biochemical analysis of subunit proteins has made it clear that the different types of amyloid are really different diseases and that amyloidosis is a generic term. Protein subunits of the AL ("primary") and AA ("secondary") types have exhibited heterogeneity of chain length and amino acid sequence that may extend to the soluble serum precursors from which they are derived. Since precursor overproduction may not always be associated with tissue deposition, proteolysis, either locally or at a site distant from the source of synthesis, may be the more crucial factor in amyloidogenesis. Availability of protein subunits, and the production of monospecific antisera to them, have opened up new diagnostic modalities and provided a more rational basis of classification for these disorders.

ACKNOWLEDGMENTS

We would like to thank Ms. Julie Lyons for her secretarial expertise.

Literature Cited

1. Franklin, E. C., Zucker-Franklin, D. 1972. Current concepts of amyloid. *Adv. Immunol.* 15:249–304
2. Glenner, G. G. 1980. Amyloid deposits and amyloidosis. *N. Engl. J. Med.* 302:1283–92, 1333–43
3. Gorevic, P. D. 1980. The amyloid diseases: clinicopathologic and biochemical correlations. In *Clinical Immunology Update,* ed. E. C. Franklin, Vol. 2, In press
4. Mandema, E. 1967. *Proc. Symp. Amyloidosis,* ed. L. Ruinen, J. H. Schulten, A. S. Cohen. Amsterdam: Excerpta Medica
5. Wegelius, O., Pasternack, A., eds. 1976. *Amyloidosis.* London: Academic
6. Glenner, G. G., ed. 1980. *Symposium on Amyloidosis.* Amsterdam: Excerpta Medica
7. Cohen, A. S. 1966. Preliminary chemical analysis of partially purified amyloid fibrils. *Lab. Invest.* 15:66–80
8. Glenner, G. G., Bladen, H. A. 1966. Purification and reconstitution of the periodic fibril and unit structure of human amyloid. *Science* 154:271–12
9. Pras, M., Schubert, M., Zucker-Franklin, D., Rimon, A., Franklin, E. C. 1968. The characterization of soluble amyloid prepared in water. *J. Clin. Invest.* 47:924–33
10. Eanes, E. D., Glenner, G. G. 1968. X-ray diffraction studies of amyloid filaments. *J. Histochem. Cytochem.* 16: 673–77
11. Termine, J. D., Eanes, E. D., Ein, D., Glenner, G. G. 1972. Infrared spectroscopy of human amyloid fibrils and immunoglobulin proteins. *Biopolymers* 11:1103–13
12. Cooper, J. H. 1974. Selective amyloid staining as a function of amyloid composition and structure. *Lab. Invest.* 31:323–28
13. Cohen, A. S., Wegelius, O. 1980. Classification of amyloid: 1979–1980. *Arthritis Rheum.* 23:644–45
14. Glenner, G. G., Keiser, H. R., Bladen, H. A., Cuatrecasas, P., Eanes, E. D., Ram, J. S., Kanfer, J. N., Delellis, R. A. 1968. Amyloid VI. Comparison of two morphologic components of human amyloid deposits. *J. Histochem. Cytochem.* 16:633–43
15. Pepys, M. B., Dash, A. C., Fletcher, T. C., Richardson, N., Munn, E. A., Feinstein, A. 1978. Analogues in other mammals and in fish of human plasma proteins, C-reactive protein and amyloid P-component. *Nature* 273:168–70
16. Cathcart, E. S., Comerford, F. R., Cohen, A. S. 1965. Immunological studies of a protein extracted from human secondary amyloid. *N. Engl. J. Med.* 273:143–46
17. Thompson, A. R., Enfield, D. L. 1978. Human plasma P-component isolation and characterization. *Biochemistry.* 17:4304–11
18. Levo, Y., Frangione, B., Franklin, E. C. 1977. Amino acid sequence similarities between amyloid P-component, C1t and CRP. *Nature* 268:56–57
19. Pepys, M. B., Dash, A. C., Markham, R. E., Thomas, H. C., Williams, B. D., Aviva, P. 1978. Comparative clinical study of protein SAP (amyloid P-component) and C-reactive protein in serum. *Clin. Exp. Immunol.* 32:119–24
20. Pepys, M. B., Dash, A. C., Munn, E. A., Feinstein, A., Skinner, M., Cohen, A. S., Gewurz, H., Osmand, A. P., Painter, R. H. 1977. Isolation of amyloid P-component (protein AP) from normal serum as a calcium-dependent binding protein. *Lancet* 1, 1029–31
21. Glenner, G. G., Terry, W. D., Isersky, C. 1973. Amyloidosis: Its nature and pathogenesis. *Semin. Hematol.* 10: 65–86
22. Glenner, G. G., Terry, W. D., Harada, M., Isersky, C., Page, D. L. 1971. Amyloid fibril proteins: Proof of homology with immunoglobulin light chains by sequence analysis. *Science* 172:1150–51
23. Kimura, S., Guyer, R., Terry, W. D., Glenner, G. G. 1972. Chemical evidence for lambda-type amyloid proteins. *J. Immunol.* 109:891–92
24. Glenner, G. G., Ein, D., Eanes, D., Bladen, H. A., Terry, W., Page, D. 1971. The creation of "amyloid" fibrils from Bence-Jones proteins in vitro. *Science* 174:712–14
25. Linke, R. P., Zucker-Franklin, D., Franklin, E. C. 1973. Morphologic, chemical and immunologic studies of amyloid-like fibrils formed from Bence-Jones proteins by proteolysis. *J. Immunol.* 111:10–23
26. Epstein, W. V., Tan, M., Wood, I. S. 1974. Formation of "amyloid" fibrils *in vitro* by action of human kidney lysosomal enzymes on Bence-Jones proteins. *J. Lab. Clin. Med.* 84:107–10
27. Isersky, C., Ein, D., Page, D. L., Harada, M., Glenner, G. G. 1972. Immunochemical cross reactions of human amyloid proteins with immunoglobulin light chains. *J. Immunol.* 108:486–93

28. Levin, M., Franklin, E. C., Frangione, B., Pras, M. 1972. The amino acid sequence of a major nonimmunoglobulin component of some amyloid fibrils. *J. Clin. Invest.* 5:2773–76
29. Gorevic, P. D., Greenwald, M., Frangione, B., Pras, M., Franklin, E. C. 1977. The amino acid sequence of duck amyloid A (AA) protein. *J. Immunol.* 118:1113–18
30. Sletten, K., Husby, G. 1974. The complete amino acid sequence of nonimmunoglobulin amyloid fibril protein AS in rheumatoid arthritis. *Eur. J. Biochem.* 41:117–25
31. Levin, M., Pras, M., Franklin, E. C. 1973. Immunologic studies of the major nonimmunoglobulin protein of amyloid I. Identification and partial characterization of a related serum component. *J. Exp. Med.* 138:373–80
32. Rosenthal, C. J., Franklin, E. C., Frangione, B., Greenspan, J. 1976. Isolation and partial characterization of SAA—an amyloid related protein from human serum. *J. Immunol.* 116:1415–18
33. Linke, R. P., Sipe, J. D., Pollock, P. S., Ignaczak, T. F., Glenner, G. G. 1975. Isolation of a low molecular weight serum component antigenically related to an amyloid fibril protein of unknown origin. *Proc. Natl. Acad. Sci. USA* 72:1473–76
34. Benditt, E. P., Eriksen, N. 1977. Amyloid protein SAA is associated with high density lipoprotein from human serum. *Proc. Natl. Acad. Sci. USA* 74:4025–28
35. Rosenthal, C. J., Franklin, E. C. 1977. Serum amyloid A (SAA) protein interaction with itself and serum albumin. *J. Immunol.* 119:630–34
36. Franklin, E. C., Prelli, F., Frangione, B. Unpublished observations
37. Costa, P. P., Figueira, A. S., Bravo, F. R. 1978. Amyloid fibril protein related to prealbumin in familial amyloidotic polyneuropathy. *Proc. Natl. Acad. Sci. USA* 75:4499–4503
38. Sletten, K., Westermark, P., Natvig, J. B. 1976. Characterization of amyloid fibril proteins from medullary carcinoma of the thyroid. *J. Exp. Med.* 143:993–98
39. Linke, R. P., Eanes, E. D., Termine, J. D., Bladen, H. A., Glenner, G. G. 1976. Formation of amyloid-like fibrils from insulin and glucagon in vitro. See Ref. 5, pp. 371–73
40. Westermark, P., Natvig, J. B., Johansson, B. 1977. Characterization of an amyloid fibril protein from senile cardiac amyloid. *J. Exp. Med.* 146:631–37
41. Gorevic, P. D., Cleveland, A. B., Wright, J. R., Zucker-Franklin, D., Frangione, B., Franklin, E. C. 1980. Ultrastructural studies and partial characterization of a low molecular weight component of senile cardiac amyloid fibrils. See Ref. 6, pp. 366–72
42. Epstein, W. V., Tan, M. 1966. Increase of L-chain proteins in the sera of patients with systematic lupus erythematosus and the synovial fluids of patients with peripheral rheumatoid arthritis. *Arthritis Rheum.* 9:713–19
43. Rosenthal, C. J., Franklin, E. C. 1975. Variation with age and disease of an amyloid A protein-related serum component. *J. Clin. Invest.* 55:746–53
44. Terry, W. D., Page, D. L., Kimura, S., Isobe, T., Osserman, E. F., Glenner, G. G. 1973. Structural identity of Bence-Jones and amyloid fibril proteins in a patient with plasma cell dyscrasia and amyloidosis. *J. Clin. Invest.* 52:1276–81
45. Lian, J. B., Skinner, M., Benson, M. D., Cohen, A. S. 1977. Fractionation of primary amyloid fibrils: Characterization and chemical interaction of the subunits *Acta Biochim. Biophys.* 171:197–205
46. Isobe, T., Osserman, E. F. 1974. Patterns of amyloidosis and their association with plasma cell dyscrasia, monoclonal immunoglobulins and Bence-Jones proteins. *N. Engl. J. Med.* 290:473–77
47. Levo, Y., Pick, A. I., Frohlichman, R. 1976. Predominance of λ-type Bence-Jones proteins in patients with both amyloidosis and plasma cell dyscrasias. See Ref. 5, pp. 291–97
48. Gorevic, P. D., Levo, Y., Frangione, B., Franklin, E. C. 1978. Polymorphism. of tissue and serum amyloid A (AA and SAA) proteins in the mouse. *J. Immunol.* 121:138–40
49. Lavie, G., Zucker-Franklin, D., Franklin, E. C. 1978. Degradation of serum amyloid A protein by surface-associated enzymes of human blood monocytes. *J. Exp. Med.* 148:1020–31
50. Lavie, G., Zucker-Franklin, D., Franklin, E. C. 1980. Elastase-type proteases on the surface of human blood monocytes: Possible role in amyloid formation. *J. Immunol.* 125:175–80
51. Sipe, J. D., McAdam, K. P. W. J., Torain, B. F., Glenner, G. G. 1976. Conformational flexibility of the serum amyloid precursor SAA. *Br. J. Exp. Pathol.* 57:582–92
52. Segrest, F. P., Pownall, H. J., Jackson, R. L., Glenner, G. G., Pollock, P. S.

1976. Amyloid A: amphipathic helixes and lipid binding. *Biochemistry* 15:3187–91

53. Kim, C. T., Shirahama, T, Cohen, A. S. 1967. The lipid content of amyloid fibrils purified by a variety of methods. *Am. J. Pathol.* 50:869–86
54. Shirahama, T., Benson, M. D., Cohen, A. S., Tanaka, A. 1973. Fibrillar assemblage of variable segments of immunoglobulin light chains: An electron microscopic study. *J. Immunol.* 110: 21–30
55. Sorenson, G. D., Binington, H. B. 1964. Resistance of murine amyloid fibril proteins to proteolytic enzymes. *Fed. Proc.* 23:550
56. Zucker-Franklin, D. 1970. Immunophagocytosis of amyloid fibrils by leukocytes. *J. Ultrastruct. Res.* 32:247–57
57. Ram, J. S., Delellis, R. A., Glenner, G. G. 1968. Amyloid IV. Is human amyloid immunogenic? *Int. Arch. Allergy Appl. Immunol.* 34:269–82
58. Deleted in proof
59. Deleted in proof
60. Kjeldsberg, C. R., Eyre, H. J., Totzke, H. 1977. Evidence for intracellular amyloid formation in myeloma. *Blood* 50:493–504
61. Baumal, R., Ackerman, A., Wilson, B. 1975. Immunoglobulin biosynthesis in myeloma-associated and casein- and endotoxin-induced murine amyloidosis. *J. Immunol.* 114:1785–91
62. Buxbaum, J. N., Hurley, M. E., Chuba, J., Spiro, T. 1979. Amyloidosis of the AL type: Response to therapy with alkylating agents and prednisone. *Am. J. Med.* 67:867–78
63. Cohen, A. S., Cathcart, E. S. 1972. Casein-induced experimental amyloidosis *Methods Achiev. Exp. Pathol.* 6:207–42
64. Scheinberg, M. A., Cathcart, E. S. 1976. Comprehensive study of humoral and cellular immune abnormalities in 26 patients with systemic amyloidosis. *Arthritis Rheum.* 19:173–81
65. Hardt, F., Ranlov, P. 1973. Amyloidogenesis. *ISR J. Med. Sci.* 9:857–61
66. Smetana, H. 1927. The relationship of the reticuloendothelial system to the formation of amyloid. *J. Exp. Med.* 45:619
67. Teilum, C. 1964. Pathogenesis of amyloidosis. The two phase cellular theory of local secretion. *Acta Pathol. Microbiol. Scand.* 61:21–45
68. Albore, S., Saavedra, G., Rose, G. G., Ibanez, M. L., Russell, W. O., Grey, C. E., Dmochowski, L. 1964. The amyloid in solid carcinoma of the thyroid gland. *Lab. Invest.* 13:77–93
69. Cohen, A. S., Gross, B. A., Shirahama, T. 1965. The light and electron microscopic autoradiographic demonstration of local amyloid formation in spleen explants. *Am. J. Pathol.* 47:1079–111
70. Bari, W. A., Pettengill, O. S., Sorenson, G. D. 1969. Electron microscopy and electron microscopic autoradiography of splenic cell cultures from mice with amyloidosis. *Lab. Invest.* 20:234–42
71. Baumal, R., Sklar, S., Wilson, B., Laskov, R. 1978. Casein-induced murine amyloidosis. Amyloidogenesis in vitro by monolayer spleen explants of casein-injected mice. *Lab. Invest.* 39:632–39
72. Hardt, F., Ranlov, P. 1976. Transfer amyloidosis. *Int. Rev. Exp. Pathol.* 16:273–334
73. Benson, M. D., Scheinberg, M. A., Shirahama, T., Cathcart, E. S., Skinner, M. 1977. Kinetics of serum amyloid protein A in casein-induced murine amyloidosis. *J. Clin. Invest.* 59:412–17
74. McAdam, K. P. W. J., Elin, R. J., Sipe, J. D., Wolff, S. M. 1978. Changes in human serum amyloid A and C-reactive protein after etiocholanolone induced inflammation. *J. Clin. Invest.* 61:390–94
75. Sipe, J. D. 1978. Induction of the acute-phase serum protein SAA requires both RNA and protein synthesis. *Br. J. Exp. Pathol.* 59:305–10
76. Sipe, J. D., Vogel, S. N., Ryan, J. L., McAdam, K. P. W. J., Rosenstreich, D. L. 1979. Detection of a mediator derived from endotoxin-stimulated macrophages that induces the acute phase serum amyloid A response in mice. *J. Exp. Med.* 150:597–606
77. Sipe, J. D., McAdam, K. P. W. J., Uchino, F. 1978. Biochemical evidence for the biphasic development of experimental amyloidosis. *Lab. Invest.* 38:110–14
78. Kisilevesky, R., Benson, M. D., Axelrad, M. A., Boudreau, L. 1979. The effect of a liver protein synthesis inhibitor on plasma SAA levels in a model of accelerated amyloid deposition. *Lab. Invest.* 41:206–10
79. Benson, M. D., Kleiner, E. 1979. Synthesis of serum amyloid protein A (SAA) by liver. *Clin. Res.* 27:321A (Abstr.)
80. Cornwell, G. G., Husby, G., Westermark, P., Natvig, J. B., Michaelsen, T., Skogen, B. 1977. Identification and characterization of different amyloid fibril proteins in tissue section. *Scand. J. Immunol.* 6:1071–80

81. Cornwell, G. G., Natvig, J. B., Westermark, P., Husby, G. 1977. Amyloid associated with senile cardiac tissue and medullary carcinoma of the thyroid: demonstration of unique fibril proteins in tissue section (abstr.) *Clin. Res.* 25:663A
82. Van Rijswijk, M. H., Van Huesden, C. W. G. J. 1979. The potassium permanganate method. A reliable method for differentiating amyloid AA from other forms of amyloid in routine practice. *Am. J. Pathol.* 97:43–54
83. Jacob, E. T., Bar-Nathan, N., Shapira, Z., Gafni, J. 1979. Renal transplantation in the amyloidosis of Familial Mediterranean Fever. *Arch. Intern. Med.* 139:1135–38
84. Fitchen, J. H. 1979. Amyloidosis and granulomatous ileocolitis. Regression after surgical removal of the involved bowel. *N. Engl. J. Med.* 292:352–53
85. Zemer, D., Pras, M., Sohar, E., Gafni, J. 1976. Colchicine in familial Mediterranean fever. *N. Engl. J. Med.* 294:170
86. Rosenthal, C. J. 1977. Serum amyloid A (SAA) effect on human B lymphocytes. *Clin. Res.* 25:366A (Abstr.)
87. Benson, M. D., Aldo-Benson, M. 1979. Effect of purified protein SAA on immune response in vitro: mechanisms of suppression. *J. Immunol.* 122:2077–87
88. Kyle, R. A., Baird, E. D. 1975. Primary systemic amyloidosis: comparison of melphalan and prednisone versus placebo. *Blood* 52:818–27
89. Osserman, E. F., Isobe, T. 1976. Effects of dimethylsulfoxide (DMSO) on Bence-Jones proteins, amyloid fibrils and casein-induced amyloidosis. See Ref. 5, pp. 247–57
90. Ravid, M. D., Kedar (Keizman), I., Sohar, E. 1977. Effect of a single dose of dimethyl-sulfoxide on renal amyloidosis. *Lancet* 1:730–31

Ann. Rev. Med. 1981. 32:273–89
Copyright © 1981 by Annual Reviews Inc. All rights reserved

PATHOPHYSIOLOGY OF PREECLAMPSIA

❖7398

Marshall D. Lindheimer, M.D. and Adrian I. Katz, M.D.

Departments of Obstetrics and Gynecology and Medicine, the University
of Chicago, Pritzker School of Medicine, Chicago, Illinois 60637

Introduction

Preeclampsia, a hypertensive disorder peculiar to pregnancy, is associated with proteinuria, edema, and at times coagulation abnormalities. It occurs primarily in nulliparas, usually after the 20th gestational week and most frequently near term. Because preeclampsia can progress rapidly to a convulsive phase termed eclampsia, it is still a major cause of fetal and maternal morbidity and mortality.

Few will contest the preceding description but the pathophysiology of this disease, and especially its management, are subjects of persistent and heated controversy. A recent mail survey in the United States was quite revealing: 86% of academic obstetricians condemned the use of diazoxide combined with diuretic administration when treating severe hypertension near term, while many nephrologists in this poll favored such therapy (1). There are similar arguments concerning the level to which maternal blood pressure should be decreased and the place of antithrombotic therapy in the management of pregnancy-induced hypertension.

Problems in Diagnosis

One of the reasons why the literature dealing with preeclampsia is confusing is that it is often impossible to distinguish clinically between preeclampsia, essential or secondary hypertension, renal disease, and combinations of these separate entities. This is best illustrated in studies where the etiology of hypertension complicating pregnancy has been diagnosed by renal biopsy (cited in 2). Table 1 summarizes the experience at our institution, where of 176 women biopsied postpartum because of hypertension complicating pregnancy only 56% had the renal lesion characteristic of preeclampsia (3). This diagnosis was most often incorrect in multiparas, and a surprisingly large number of women had unsuspected parenchymal renal disease. In

0066-4219/81/0401-0273$01.00

another study, in which an academic obstetrician and a nephrologist recorded their impressions prior to biopsy, an exact diagnosis was made on clinical grounds in only 58% of the patients (4). However, renal biopsies are performed infrequently and investigators use other criteria to increase the accuracy of diagnosis, such as studying primigravid populations only, requiring that patients manifest both proteinuria and hypertension, and relying on an increment in serum urate levels. Still, proteinuria may increase to the pathologic range during pregnancy in hypertensive women with nephrosclerosis or other renal parenchymal lesions (5), and serum urate levels can increase near term in normal pregnancy as well as in women with intrauterine growth retardation (6, 7).

Another factor confounding the differential diagnosis of the hypertensive disorders in gestation is related to the physiological change in blood pressure that occurs during pregnancy. The diastolic blood pressure decreases early in normal gestation, and by midterm levels are about 10 mm Hg lower than those in nonpregnant patients (2, 8, 9). Blood pressure increases thereafter, slowly approaching prepregnant values just prior to delivery. In some women with essential hypertension, mean blood pressure decreases to seemingly normal levels early in pregnancy. These women are then erroneously labeled preeclamptic when frankly hypertensive values are recorded near term.

In summary, there are many problems in interpreting studies of preeclamptic (or "toxemic") populations in which the diagnosis is based on clinical criteria alone, especially in series where many of the patients are multiparas. More reliable are reports restricted to primigravid women with hyperuricemia and proteinuria and whose midtrimester diastolic blood pressure was less than 70 mm Hg. With this background in mind we review current concepts of the pathophysiology of preeclampsia. With few excep-

Table 1 Renal pathology in 176 hypertensive gravidas[a]

Diagnosis	Number of patients	Primigravidas	Multiparas
Preeclampsia[b]	96	79	17
with nephrosclerosis	13	6	7
with renal disease	3	1	2
with both	2	1	1
Nephrosclerosis	19	3	16
with renal disease	4	2	2
Renal disease	31	12	19
Normal histology	8	0	8

[a] From Reference 3.

[b] Only glomerular endotheliosis on biopsy.

tions, citations focus on studies reported after 1975; for reviews of the earlier literature, see Chesley (10) and Lindheimer & Katz (2, 11).

Hypertension in Preeclampsia

The blood pressure rise in preeclampsia usually occurs in late pregnancy[1]. However, midterm pressures, though still normal, are in fact higher in many (but not all) women who later manifest hypertension near term, compared to gravidas who will remain normotensive throughout gestation (9, 12–14). Studies of remote hypertension in such patients suggest the following: Gravidas manifesting "normal" midterm diastolic pressure of 70 mm Hg or greater often become frankly hypertensive in late pregnancy and belong to two populations; women who are actual or potential chronic hypertensives, and gravidas who are destined to manifest preeclampsia and whose peripheral vascular tone may be inappropriately elevated quite early in gestation (12, 15).

The severity of hypertension may be a clue to the underlying disease. Chesley (15) noted that if systolic blood pressure exceeded 160 mm Hg during eclampsia, especially if the disease appeared prior to the 32nd gestational week, the patient often manifested essential hypertension later in life. In our study, in which diseases were classified by morphologic criteria, the blood pressure averaged 151/99 mm Hg in women with nephrosclerosis, compared to 145/95 mm Hg in patients with pure preeclampsia (glomerular endotheliosis only on biopsy) and 135/87 mm Hg in gravidas with renal disease (3). While all these values were significantly different from each other, the range of pressures was wide enough to preclude use of the blood pressure level as a diagnostic criterion.

Hypertension in preeclampsia shows an interesting reversal of the circadian blood pressure rhythms that normally occur in pregnant and nonpregnant populations, i.e. morning peaks and nighttime nadirs. In preeclampsia the rhythm may be abolished or even reversed, with the highest pressures occuring at night (16–18).

The rise in blood pressure in preeclampsia is typically labile because of an intense vascular sensitivity to endogenous pressor peptides and amines (2, 16–19). Whereas normal pregnant women are quite resistant to the pressor effects of infused angiotensin, those destined to develop preeclampsia manifest increased pressor responsiveness to this peptide many weeks prior to any clinical signs of the disease (20). In women with chronic essential hypertension, sensitivity to infused angiotensin also increases prior to the development of superimposed preeclampsia (21). In this respect there are discrepancies in the literature dealing with the renin-angiotensin system in preeclampsia. Initially, investigators found that renin concentration and activity and angiotensin II levels were decreased in this disease compared to normal pregnancy (where all components of the renin-angiotensin-aldosterone system are elevated). More recently, however, several authors have noted increments in both plasma renin activity and circulating angiotensin II in preeclampsia (2, 19, 22–25). One must keep in mind, though, that even at lower concentrations these pressor substances probably play a pathogenetic role, if the exquisite sensitivity of the vasculature is taken into account. Some observations, however, argue against a role of angiotensin II in the pathogenesis of hypertension in preeclampsia, e.g. saralasin (E. M. Symonds, personal communication) or converting enzyme inhibitor (SQ 20881) (26) was ineffective when given shortly after delivery to preeclamptic patients whose blood pressure had remained elevated.

---

[1] There are rare instances when preeclampsia-eclampsia occurs prior to or shortly after midterm. These include patients with hydatiform mole, fetal hydrops, underlying renal parenchymal disease, and multiple fetuses.

Other humoral substances implicated in the pathogenesis or maintenance of the hypertension in preeclampsia include catecholamines, prolactin, vasopressin, prostaglandins, and "unknown" pressor substances. Norepinephrine and epinephrine levels may rise and the sensitivity to their pressor effects is also increased in preeclampsia (2, 19). Levels of dopamine-beta hydroxylase appear unaltered (25). Mendlowitz (27), who in 1961 noted that digital reactivity to norepinephrine is increased in mid-pregnancy in women who later developed "toxemia," has recently espoused a hypothesis in which the role of catecholamines in the increased blood pressure during preeclampsia assumes new importance. There is a rapidly growing literature describing (in nonpregnant hypertensives) membrane transport abnormalities characterized by increased intracellular calcium in smooth muscle, which sensitizes it to constrictor stimuli. Thus, Mendlowitz suggests that preeclampsia occurs in gravidas who manifest a temporary defect in their calcium transport system. In this formulation, gestation is an anabolic state in which some women fail to produce sufficient calcium-transporting protein. A report by Kuhnert et al (28), who noted decrements in Na-K-ATPase of cord blood erythrocytes from preeclamptic women, is of interest, since a defect in the sodium pump has been postulated in the pathogenesis of the cellular calcium transport defect discussed above.

Other hormones implicated in preeclampsia include prolactin and vasopressin. Prolactin was said to antagonize the effect of circulating pressor substances (29), but levels of this hormone in preeclamptics vary, having been reported as decreased, normal, or elevated (30–32). Vasopressin, suspected almost 50 years ago of playing a role in "toxemia," has recently been implicated in the blood pressure rise seen in several experimental animal models of hypertension. Its role in preeclampsia is uncertain, but in one report levels of antidiuretic hormone measured by radioimmunoassay were lower in preeclamptics than in normal gravidas (33). The literature also abounds with references to "unknown humoral substances" that are said to play important roles in the pathophysiology of preeclampsia (16, 19, 34, 35). Plasma (or plasma dialysates) from preeclamptic women has vasoconstrictor properties and decreases the short circuit current (a measure of sodium transport) in toad bladder preparations. Many of these reports await confirmation and further clarification.

Most recently, investigators have turned their attention to the role of vasodilating prostaglandins (of the E and I series). They propose that impaired ability of the uterus or blood vessels to synthesize these lipid hormones may cause hypertension because of an imbalance between circulating pressor substances and the counteracting effect of prostaglandins on smooth muscle. Said otherwise, increased production of uterine or vessel wall prostaglandin E or I is required to counteract the vasoconstrictive effects of the elevated angiotensin II concentrations characteristic of normal pregnancy. Experiments in animal models reveal that the pregnant uterus is a rich source of prostaglandin E, some of which reaches the circulation, and that inhibition of PGE synthesis increases the blood pressure (36, 37). Increments in urinary prostaglandin excretion have been described in normal human pregnancy, but data on circulating levels are conflicting (38, 39). Of interest are observations that ingestion of aspirin or indomethacin, both cyclooxygenase inhibitors, decreases the resistance to infused angiotensin normally present in gestation to the level seen in nonpregnant subjects or in women with preeclampsia (40), and that prostaglandin synthesis is impaired in placentas from preeclamptic patients (2, 22, 41).

It was suggested that production of progesterone, a vasodilating hormone with natriuretic actions, may be decreased in preeclampsia, but Weinberger et al (42) found that circulating levels of progesterone were unaltered in this condition. Infusions of the progesterone metabolite 5-hydroxyprogesterone into women with preeclampsia did abolish their increased sensitivity to angiotensin II, so that their response to this substance became similar to that of normal gravidas (43). The same authors noted a similar effect in preeclamptics given theophylline, and postulated a role for cyclic nucleotides in the abnormal vascular reactivity that accompanies this disease (44). Several excellent reviews of factors controlling vascular responsiveness in normal and hypertensive pregnancies have been recently published (45, 46).

Another interesting facet of the altered vascular reactivity in preeclamptic women is the development of a pressor response to supine posture several weeks prior to frank hypertension. Gant et al (47) demonstrated that in nulliparous normotensive women who subsequently developed preeclampsia, diastolic pressure increased 20 mm Hg or more when they were turned from the lateral recumbent to a supine position. The reliability of this maneuver, called the "supine pressor response test" or "roll over test," is currently being debated (48).

Altered Renal Function and Proteinuria

Preeclampsia is accompanied by a morphologic lesion termed glomerular endotheliosis (19, 49). This lesion is considered by most, but not all (50), investigators to be virtually pathognomonic of the disease. The morphologic lesion is accompanied by functional changes in renal hemodynamics and the tubular handling of urate, as well as by abnormal protein excretion.

Both the glomerular filtration rate (GFR) and renal plasma flow decrease in preeclampsia, and since the decrement in GFR is greater, the filtration fraction falls (19, 51). The drop in GFR is approximately 25% in mild cases, whereas renal hemodynamics increase 30–50% in normal pregnancy. Despite morphological evidence of glomerular cell swelling, ischemia, and obliteration of the urinary space, GFR in preeclamptic women often remains above prepregnancy values. Therefore, levels for serum creatinine and urea nitrogen of 0.9 mg and 15 mg per 100 ml respectively, which are considered normal in nonpregnant subjects, indicate decreased function in gravidas.

Uric acid clearance decreases in preeclampsia (2, 19, 51). This decrement may occur earlier (2, 52) and be more profound than that in GFR, so that the urate/inulin clearance ratio is also decreased. The level of hyperuricemia correlates directly with the decrement in plasma volume that occurs in preeclampsia, and indirectly with the plasma renin activity (7). High urate levels also correlate with the severity of the preeclamptic renal lesion (4) as well as with poor fetal outcome (2, 19, 53).

Abnormal proteinuria almost always accompanies preeclampsia and the diagnosis is suspect without this sign, even though glomerular endotheliosis has sometimes been described in the absence of increased protein excretion (54). Proteinuria may be minimal, moderate, or severe, i.e. in the nephrotic range. The occurrence of nephrotic syndrome deserves emphasis, for in the past heavy proteinuria was believed to be uncommon in preeclampsia and when it occurred it was considered an indication of a severe form of the disease. It is now apparent that preeclampsia is the most common cause of nephrotic syndrome in pregnancy (3, 55–57) and that this diagnosis may be overlooked if one fails to consider diastolic levels between 80–90 mm Hg as abnormal in these patients. In our studies the severity of maternal disease was similar in preeclamptic women with heavy proteinuria and those excreting less than 3.5 g per 24 h, although a small but significant increase in fetal loss occurred in the women with severe proteinuria (3, 55).

Proteinuria in preeclampsia is nonselective (58), even though the lesion is completely reversible (19). Proteinuria has been attributed to vasospasm or to the glomerular lesion that accompanies the disease. The latter is supported by the fact that the magnitude of the proteinuria correlates with the severity of the morphologic lesion (3).

Sodium Retention and Edema

While fluid retention and edema occur in preeclampsia, they are also a feature of normal pregnancy. Plasma volume increases markedly in uncomplicated gestation and 60–80% of normal gravidas manifest dependent edema sometime during their pregnancy, which on occasion may become generalized (2, 59, 60). Edema alone imparts no adverse prognosis to the reproductive outcome (61); fetal size may even be greater in such patients (59, 60). Thus, sodium is no longer considered undesirable in pregnancy or causally linked to preeclampsia, and prophylactic salt restriction and/or diuretic therapy of normotensive gravidas is no longer condoned. Nevertheless, rapid weight gain and generalized edema is a feature of preeclampsia. The ability to excrete sodium is usually impaired in preeclampsia (2, 19, 60), but the disease can occur without fluid retention, in which case it is often more severe (60). However, even when edema is present, plasma volume is decreased (compared with normal pregnancy) and hemoconcentration occurs. Furthermore, while preeclamptics excrete sodium poorly, they may paradoxically display an exaggerated natriuresis when challenged with salt later in life (62).

The cause of sodium retention in certain women with preeclampsia is obscure. The decrement in GFR seems insufficient explanation, and aldosterone levels in early and midpregnancy are similar in women destined to develop third-trimester hypertension and in those who remain normotensive through gestation (63). While concentrations of aldosterone are markedly increased in normal gravidas, the levels actually decrease in preeclamptics (2, 19, 22, 33, 39, 42, 59).

The metabolism of other mineralocorticoids that play a role in sodium retention has barely been studied. Desoxycorticosterone (DOC), whose levels increase in pregnancy (2, 59, 64), may be largely of fetal origin. Since in pregnancy concentrations of DOC are not suppressible by physiological maneuvers such as sodium loading (64), this mineralocorticoid could play a role in maintaining edema in preeclampsia.

The decrease in plasma volume accompanied by salt retention in preeclampsia is the basis of a major controversy. Some investigators believe that decreased intravascular volume is the primary event in preeclampsia, which may be responsible for the rise in blood pressure (i.e. placental hypoperfusion may induce release of a pressor substance from the uterus). This naturally leads to the recommendation of volume expansion therapy, an approach rejected by traditional hypertension experts. The latter believe that decrements in volume are secondary to vasoconstriction, and since cardiac output is often unchanged or increased in preeclampsia (2, 16, 65) while myocardial performance is compromised, these authorities warn that treatment with volume expanders is dangerous and may lead to aggravation of the hypertension as well as to pulmonary edema; they advocate instead use of vasodilating drugs and diuretics. Data in the literature on the effects of plasma expansion are equivocal (66), but recent observations seem to provide theoretical support for this approach. In preeclampsia, the decrease in plasma volume precedes the occurrence of hypertension (67), and the central venous pressure is low and varies inversely with the severity of the blood pressure elevation (68).

Coagulation Abnormalities

There is considerable controversy concerning the incidence and significance of coagulation abnormalities in preeclampsia. Many reports describe varying degrees of disseminated intravascular coagulation in preeclamptic women, and some suggest that coagulopathies play a role in the pathogenesis of the disease (2, 22, 69). Redman et al (70) state that a decrement in platelet count is the earliest sign of preeclampsia and occurs weeks before clinical signs are evident. Abnormal platelet localization in the kidney is said to persist postpartum despite normalization of the blood pressure (71). It should be emphasized that although laboratory evidence of coagulation is easily demonstrated in patients with severe or fulminating disease, specialized testing (e.g. fibrin chromatography, the relationship of factor VIII to its antigen, or platelet survival times) is required to demonstrate intravascular coagulation when preeclampsia is mild (2, 19, 69, 72–75). Nevertheless, cases in which signs of coagulopathy predominate while hypertension is only mild or moderate are occasionally seen. In addition, some investigators (cited in 2, 20, 49, 76) state that fibrin or fibrin-like material is invariably present in renal biopsies demonstrating glomerular endotheliosis.

Contrasting with the above views are studies that do not ascribe a primary role to coagulation abnormalities in the pathogenesis of preeclampsia. Pritchard and colleagues (77) performed a battery of five tests (platelet count, plasma fibrinogen and its degradation products, fibrin monomer, and thrombin time) and found evidence of disordered coagulation in only a minority of their eclamptic patients.[2] They note that the thrombocytopenia sometimes observed in preeclampsia could be due to platelet adherence at sites of disrupted vascular endothelium rather than to intravascular coagulation induced by factors postulated as initiators of the disease (e.g. thromboplastin from decidua or placenta). They further stress that in abruptio placentae, an obstetrical syndrome that may be accompanied by disseminated intravascular coagulation, hypertension, convulsions, and other manifestations seen in preeclampsia are usually absent. Decrements in the level of clotting factors in preeclampsia could also be due to liver involvement in this disease. Finally, in our renal biopsy series, immunofluorescent positive fibrin-like material was present in only 20 of 45 specimens with ultrastructural evidence of glomerular endotheliosis, and in only 8 instances did its intensity exceed one plus; glomerular fluorescence for antihemophilic globulin was present in only 5 of 10 specimens analyzed, and exceeded one plus but once (3).

---

[2] In the absence of a renal biopsy, convulsion in a hypertensive primipara is a good indicator that the patient had the morphological lesion of preeclampsia.

Disagreements over the role of coagulopathy in preeclampsia are at the root of certain controversies concerning the management of this disease. Some authors recommend therapy with antithrombotic agents such as aspirin, dipyridamole, and heparin (72, 78. 79). Others, including ourselves, counsel against this because of the risks of anticoagulation in patients with frank hypertension whose disease is sometimes complicated by liver or cerebral hemorrhages, especially since other therapeutic approaches are successful.

Eclampsia

Convulsions are the most dramatic and life-threatening complication of preeclampsia. They are characterized by a series of generalized muscular contractions that last 20–30 seconds and may recur for several minutes. After this, coma often ensues, the convulsions may reappear, and unless treatment is promptly started the patient may die. At one time maternal mortality in eclampsia averaged 20–30%, with cerebral hemorrhage as its leading cause (80). Convulsions are rarely preceded by an aura and may occur without any warning, but among the premonitory signs which should alert the clinician that this complication is imminent are nervousness, apprehension, severe headaches, visual disturbances, sudden facial edema, rapidly rising diastolic blood pressure, increasing proteinuria, decreasing urine output, and severe hemoconcentration. Experienced obstetricians initiate prophylactic anticonvulsive therapy ($MgSO_4$) even when indications are equivocal, an approach that has virtually eradicated maternal mortality in this disease.

There is a mistaken tendency to equate eclampsia with hypertensive encephalopathy. Although eclampsia correlates with the severity of the hypertension, it may also arise when blood pressure elevations are mild. For example, in 22% of the eclamptics originally described by Dieckmann (cited in 11), systolic levels never exceeded 140 mm Hg; similar observations were reported by Sheehan & Lynch (81) in their large autopsy series. Thus there is no place for ambulatory care of preeclamptics: Patients with seemingly mild disease, for example a teenage gravida with a blood pressure of 140/90 mm Hg and minimal proteinuria, have been known to convulse. *De novo* hypertension in a third-trimester nullipara is always a serious disease requiring a tentative diagnosis of preeclampsia even if no other signs are present. Suspicion of the disease alone is sufficient reason to hospitalize the patient and treat her as a potential eclamptic.

The pathogenesis of the eclamptic convulsion is poorly understood and has been scarcely investigated. Based on autopsy results, McKay (82) attributes convulsions to disseminated intravascular coagulation characterized by platelet-fibrin clots that obstruct the cerebral microcirculation. Sheehan & Lynch (80), who performed most of their autopsies between 15 minutes and 2 hours after death, do not support this view and ascribe the cerebral changes to intense vasoconstriction. There is also a high percentage of frank cerebral bleeding in autopsy material from women dying of eclampsia. According to Sheehan & Lynch most of these episodes occur after the initial convulsion.

Cerebral blood flow in preeclampsia-eclampsia is decreased (16), an observation that should be taken into account when prescribing anesthetics or sedation. There is a host of ocular signs associated with preeclampsia, including the development of amaurosis (83, 84). In one such case at our institution, temporary blindness was associated with changes consistent with occipital lobe edema as determined by computerized axial tomography. The edema reverted to normal when vision returned.

Placental Perfusion

Uteroplacental blood flow decreases in preeclampsia (16), which is a major reason for the high incidence of fetal loss, intrauterine growth retardation, and small-for-date infants. Gant and colleagues, who use the metabolic and placental clearance of dehydroisoandrosterone sulfate (DS) as an index of placental perfusion, suggest that uteroplacental blood flow may actually be higher in women destined to develop preeclampsia than in those who remain normotensive, and that perfusion starts to decrease prior to the onset of hypertension (summarized in 85). The early increment in flow is reminiscent of a view that preeclampsia occurs more readily in women with excessive placental growth ("hyperplacentosis"), while the decrease in perfusion is compatible with the theory that uterine ischemia (relative or absolute) causes the high blood pressure in this disease. It should be noted, however, that the use of DS clearance as an index of uterine perfusion has been criticized (86).

The decrease in uterine blood flow is another reason for some of the current controversy in the management of preeclampsia. Data from morphologic studies of human placentae demonstrate trophoblastic invasion of the muscular coats that surround the uterine spiral arteries. Observations in unanesthetized sheep or primates suggest that uterine arteries behave like rigid conduits, or are normally in a maximally dilated state (2, 16, 19, 85, 87, 88). Thus, believing that reductions in maternal blood pressure decrease uteroplacental perfusion, some authors caution against large decrements in mean arterial pressure in acute emergencies, especially since placental perfusion is already compromised in preeclampsia. Other investigators, whose data are derived from studies in unanesthetized rabbits, report that autoregulation of uterine blood flow is complete over a wide range of pressures in the dam; they recommend aggressive lowering of blood pressure when hypertension arises suddenly near term (36, 90, 91). Data from human pregnancy are indirect and limited, but suggest that acute reductions in maternal pressure may decrease uteroplacental perfusion (85,) If one assumes that autoregulation exists, a critical but unanswered question is how quickly it occurs, for the fetus will be compromised if ischemia persists for any length of time. There are documented instances where precipitous reductions of the blood pressure were accompanied by signs of fetal distress even when diastolic levels exceeded 85 mm Hg (91, 92). We therefore recommend an intermediate approach in which antihypertensive agents are not prescribed as long as maternal blood pressure is only mildly elevated (105 mm Hg diastolic or less); when indicated these agents should be used to decrease gently the blood pressure to levels between 90 and 100 mm Hg, which should harm neither mother nor child.

Hypothesis

Preeclampsia has been called the disease of theories. Some of these have already been mentioned in the preceding sections, but a detailed discussion of all the hypotheses regarding the etiology and pathogenesis of preeclampsia is beyond the scope of this article. Among some of the theories currently popular, we mention briefly those that attribute a major role to decreased uteroplacental perfusion, to immunological abnormalities, and to a genetic disorder. Supporting the first theory are a series of animal experiments in which acute reductions in uteroplacental blood flow were accompanied by decreases in the uterine venous efflux of prostaglandin-like materials and increases in maternal blood pressure (34, 37, 89). Also of interest are a series of reports by Abitbol and colleagues (93–95) in which aortic constriction caused hypertension in pregnant rabbits, dogs, and primates, and was accompanied by renal lesions resembling those of human preeclampsia. Certainly, development of a convincing animal model of preeclampsia would be an important step toward understanding this disease.

Recent interest has refocussed on an immunologic cause of preeclampsia. Hypersensitivity-like reactions (e.g. maternal-fetal incompatibility) were suggested and rejected since preeclampsia is primarily a disease of the first pregnancy and usually does not occur in subsequent gestations. A growing body of evidence suggests that preeclampsia may represent a failure of the maternal immune response, rather than a hypersensitivity state (96–98). Studies demonstrating significantly more sharing of fetal-maternal antigens than predicted for the general population support this view (96). Excellent reviews of observations pertinent to the hypothesis described above, as well as discussion of the mechanism whereby excessive fetomaternal immunologic compatibility may lead to preeclampsia, have been published (99–101).

Concerning genetic theories, analysis of data derived by Chesley et al (102) from studies of eclampsia in mothers and daughters supports the notion that this disease is inherited as a recessive Mendelian trait (Table 2) (103, 104). The immunologic and genetic theories may be related (100); Redman et al (105) noted a significant excess of apparent HLA homozygosity in preeclamptics and suggested that homozygosity for recessive maternal immune response genes linked to the HLA locus might predispose to severe preeclampsia.

Conclusions

Preeclampsia-eclampsia remains a major complication of pregnancy, jeopardizing both mother and fetus. There are many gaps in our knowledge concerning the cause and pathophysiology of this disease, but the little we know does allow a rational approach to management.

Preeclampsia is an explosive disease. Suspicion of this disorder requires hospitalization because the convulsive phase of the disorder (eclampsia) can occur rapidly, even when the blood pressure is only mildly elevated. The

Table 2 Preeclampsia or eclampsia in relatives of eclamptics[a]

		Preeclampsia or eclampsia during 1st pregnancy	
Relationship	Number	Expected[b]	Observed
Sisters	147	58.0	55
Daughters	187	48.9	48
1 daughter	72	18.4	18
2 daughters	28	12.5	13
3 to 5 daughters	12	8.3	8
Daughters-in-law (controls)	75	4.9	6

[a] From Reference 104.

[b] Predicted numbers if caused by single recessive gene.

pathophysiology of preeclampsia is characterized by extreme vasoconstriction, enhanced vascular reactivity to pressor peptides and amines, and decreased intravascular volume (usually despite an increased interstitial space). Other manifestations include decreased glomerular filtration rate and renal plasma flow, a marked decrement in the clearance of urate, and nonselective proteinuria, which may be in the nephrotic range. Some women with preeclampsia manifest variable signs of coagulopathy, which range from mild abnormalities in coagulation factors to severe disseminated intravascular coagulation. Marked vasoconstriction decreases cerebral perfusion and may play a role in the pathogenesis of the convulsion, but thrombosis in the cerebral microcirculation has also been implicated in the genesis of eclampsia. There are also decrements in uteroplacental perfusion, a factor that contributes to the increased fetal loss and morbidity associated with preeclampsia.

The pathogenesis of the diverse manifestations of preeclampsia is also unclear, which explains some of the controversies surrounding the current approach to management of the disease. Nevertheless, prompt hospitalization of suspected preeclamptics and therapeutic intervention aimed at reducing blood pressure and preventing convulsions have brought about a gratifying improvement in the mother's prognosis and in the outcome of her pregnancy.

ACKNOWLEDGMENTS

We thank Mrs. C. Regovic for her secretarial assistance. Some of the authors' work cited was supported by the National Institutes of Health (RR55, HD5572) and the American Heart Association.

Literature Cited

1. Lewis, P. J., Boulpitt, C. J., Zuspan, F. B. 1981. A comparison of current British and American practice in the management of hypertension in pregnancy. *J. Obstet. Gynecol.* In press
2. Lindheimer, M. D., Katz, A. I. 1981. The renal response to pregnancy. In *The Kidney,* ed. B. M. Brenner, F. C. Rector Jr. Philadelphia: Saunders. 2nd ed. In press
3. Lindheimer, M. D., Fisher, K. A., Spargo, B. H., Katz, A. I. 1981. Hypertension in pregnancy: A biopsy study with long term follow-up. *Contrib. Nephrol.* In press
4. Pollak, V. E., Nettles, J. B. 1960. The kidney in toxemia of pregnancy: A clinical and pathologic study based on renal biopsies. *Medicine* 39:469–526
5. Katz, A. I., Davison, J. M., Hayslett, J. P., Singson, E., Lindheimer, M. D. 1980. Pregnancy in women with kidney disease: *Kidney Int.* 18:192–206
6. Dunlop, W., Davison, J. M. 1977. The effect of normal pregnancy upon the renal handling of uric acid. *Br. J. Obstet. Gynaecol.* 84:13–21
7. Beaufils, M., Uzan, S., Don Simoni, R., The metabolism of uric acid in normal and pathological pregnancy: A reappraisal. *Contrib. Nephrol.* In press
8. MacGillivray, I., Rose, G. A., Rowe, B. 1969. Blood pressure survey in pregnancy. *Clin. Sci.* 37:395–407
9. Page, E. W., Christianson, R. 1976. The impact of mean arterial blood pressure in the middle trimester upon the outcome of pregnancy. *Am. J. Obstet. Gynecol.* 125:740–45

10. Chesley, L. C. 1978. *Hypertensive Disorders in Pregnancy.* New York: Appleton-Century-Crofts. 525 pp.
11. Lindheimer, M. D., Katz, A. I. 1977. *Kidney Function and Disease in Pregnancy.* Philadelphia: Lea & Febiger. 246 pp.
12. See Ref. 10, pp. 287–88
13. Fallis, N., Langford, H. G. 1963. Relation of second trimester blood pressure to toxemia of pregnancy in the primigravid patient. *Am. J. Obstet. Gynecol.* 87:123–25
14. Gallery, E. D. M., Ross, M., Hunyor, S. N., Györy, A. Z. 1977. Predicting the development of pregnancy-associated hypertension. The place of standardized measurement. *Lancet* 1:1273–75
15. See Ref. 10, pp. 421–43
16. See Ref. 10, pp. 119–53
17. Redman, C. W. G., Beilin, L. J., Bonnar, J. 1976. Variability of blood pressure in normal and abnormal pregnancy. *Perspect. Nephrol. Hypertens.* 5:53–59
18. Murnaghan, G. A., Mitchell, R. H., Ruff, S. 1980. Circadian variation in blood pressure in pregnancy. In *Pregnancy Hypertension,* ed. J. Bonnar, I. McGillivray, E. M. Symonds, pp. 107–11. Lancaster, England: MTP. 597 pp.
19. See Ref. 11, pp. 188–221
20. Gant, N. F., Daley, G. L., Chand, S., Whalley, P. J., MacDonald, P. C. 1973. A prospective study of angiotensin II pressor response throughout primigravid pregancy. *J. Clin. Invest.* 52: 2682–89
21. Gant, N. F., Jiminez, J. M., Whalley, P. J., Chand, S., MacDonald, P. C. 1977. A prospective study of angiotensin II pressor responsiveness in pregnancies complicated by chronic essential hypertension. *Am. J. Obstet. Gynecol.* 127:369–75
22. See Ref. 10, pp. 229–55
23. Broughton-Pipkin, F., Symonds, E. 1977. Factors affecting angiotensin II concentrations in the human infant at birth. *Clin. Sci. Mol. Med.* 52:449–56
24. Symonds, E. M., Broughton-Pipkin, F. 1978. Pregnancy hypertension, parity and the renin-angiotensin. *Am. J. Obstet. Gynecol.* 132:473–79
25. Annat, G., Raudrant, D., Chappe, J., Vincent, M., Thoulon, J., Dummont, M., Sassard, J. 1978. Maternal fetal plasma renin activities and dopamine B-hydroxylase activities in toxemic pregnancy. *Obstet. Gynecol.* 52:219–24
26. Sullivan, J. M., Palmer, E. J., Schoeneberger, A. A., Jenings, J. C., Morrison, J. C., Ratts, T. E. 1978. SQ 20881: Effect on eclamptic-preeclamptic women with postpartum hypertension. *Am. J. Obstet. Gynecol.* 131: 707–15
27. Mendlowitz, M. 1980. Toxemia of pregnancy and eclampsia. *Obstet. Gynecol. Survey* 35:327–31
28. Kuhnert, B. R., Kuhnert, P. M., Murray, B. A., Sokol, R. J. 1977. Na/K and Mg-ATPase activity in the placenta and in maternal and cord erythrocytes of preeclamptic patients. *Am. J. Obstet. Gynecol.* 127:56–60
29. Mati, J. K. G., Mugambi, M., Odipo, W. S., Nglui, K. 1976. Prolactin and hypertension. *Am. J. Obstet. Gynecol.* 127:616–19
30. Redman, C. W. G., Bonnar, J., Beilin, L. J., McNeilly, A. S. 1975. Prolactin in hypertensive pregnancy. *Br. Med. J.* 1:304–6
31. Dubowitz, L., Strong, F., Hawkins, D., Blair, C. M., Mashiter, K. 1975. Toxemia of pregnancy and plasma prolactin. *Br. Med. J.* 2:445
32. Yuen, B. H., Cannon, W., Wooley, S., Charles, E. 1978. Maternal plasma and amniotic fluid prolactin levels in normal and hypertensive pregnancy. *Br. J. Obstet. Gynaecol.* 85:293–98
33. Weir, R. J., Doig, A., Fraser, R., Morton, J. J., Parboosingh, J., Robertson, J. I. S., Wilson, A. 1976. Studies of the renin-angiotensin-aldosterone system: Cortisol, DOC, and ADH in normal and hypertensive pregnancy. *Perspect. Nephrol. Hypertens.* 5:251–61
34. See Ref. 10, pp. 445–76
35. Mouad-Filho, F., Salles Meirelles, R. 1978. Effects of maternal and fetal plasma from toxemic patients on the vasomotility of the human umbilical artery. *Am. J. Obstet. Gynecol.* 130: 718–21
36. Ferris, T. F., Venuto, R. C., Bay, W. H. 1976. Studies of the uterine circulation in the pregnant rabbit. *Perspect. Nephrol. Hypertens.* 5:351–60
37 Terragno, N. A., Terragno, D. A., McGiff, J. C. 1976. The role of prostaglandins in the control of uterine blood flow. *Perspect. Nephrol. Hypertens.* 5:392–97
38. Whalen, J. B. Clancey, C. J., Fairley, D., Van Orden, D. E. 1978. Plasma prostaglandins in pregnancy. *Obstet. Gynecol.* 51:52–55
39. Bay, W. H., Ferris, T. F. 1979. Factors controlling plasma renin and aldosterone in pregnancy. *Hypertension* 1:410–15

40. Everett, R. B., Worley, R. J., MacDonald, D. C., Gant, N. F. 1978. Effect of prostaglandin synthetase inhibitors on pressor response to angiotensin II in human pregnancy. *J. Clin. Endocrinol. Metab.* 46:1007–10
41. Demers, L. M., Gabbe, S. G. 1976. Placental prostaglandin levels in preeclampsia. *Am. J. Obstet. Gynecol.* 126:137–39
42. Weinberger, M. H., Kramer, N. J., Petersen, L. P., Cleary, R. E., Young, P. C. M. 1976. Sequential changes in the renin-angiotensin-aldosterone systems and plasma progesterone concentration in normal and abnormal human pregnancy. *Perspect. Nephrol. Hypertens.* 5:263–68
43. Everett, R. B., Worley, R. J., MacDonald, P. C., Gant, N. F. 1978. Modification of vascular responsiveness to angiotensin II in pregnant women by intravenously infused 5a-dihydroprogesterone. *Am. J. Obstet. Gynecol.* 131:352–57
44. Everett, R. B., Worley, R. J., MacDonald, P. C., Gant, N. F., 1978. Oral administration of theophylline to modify pressor responsiveness to angiotensin II in women with pregnancy-induced hypertension. *Am. J. Obstet. Gynecol.* 131:359–62
45. Gant, N. F., Worley, R. J. 1980. *Hypertension in Pregnancy: Concepts and Management,* pp. 11–36, New York: Appleton-Century-Crofts. 211 pp.
46. Gant, N. F., Worley, R. J., Everett, R. B., MacDonald, P. C. 1980. Control of vascular responsiveness during human pregnancy. *Kidney Int.* 18:253–58
47. Gant, N. F., Chand, S., Worley, R. J., Whalley, P. J., Crosby, V. D., MacDonald, P. C. 1974. A clinical test useful for predicting the development of acute hypertension in pregnancy. *Am. J. Obstet. Gynecol.* 120:1–7
48. Kassar, N. S., Aldridge, J., Quirk, B. 1980. Roll over test. *Obstet. Gynecol.* 55:411–13
49. Sheehan, H. L. 1980. Renal morphology in preeclampsia. *Kidney Int.* 18:241–52
50. Robson, J. S. 1976. Proteinuria and the renal lesion in preeclampsia and abruptio placentae. *Perspect. Nephrol. Hypertens.* 5:61–72
51. See Ref. 10, pp. 154–97
52. Gallery, E. D. M., Györy, A. Z. 1979. Glomerular and proximal renal tubular function in pregnancy-associated hypertension: a prospective study. *Eur. J. Obstet. Gynecol. Reprod. Biol.* 9:3–12
53. Redman, C. W. G., Beilin, L. J., Bonnar, J., Wilkinson, R. H. 1976. Plasma-urate measurements in predicting fetal death in hypertensive pregnancy. *Lancet* 1:1370–73
54. Nochy, D., Birembout, P., Hinglais, N., Freund, M., Idatte, J. M., Jacquot, C., Chartier, M., Bariety, J. 1980. Renal lesions in the hypertensive syndromes of pregnancy: immunomorphological and ultrastructural studies in 114 cases. *Clin. Nephrol.* 13:155–62
55. Fisher, K., Ahuja, S., Luger, A., Spargo, B. H., Lindheimer, M. D. 1977. Nephrotic proteinuria with preeclampsia. *Am. J. Obstet. Gynecol.* 129:643–46
56. First, M. R., Ooi, B. S., Wellington, J., Pollak, V. E. 1978. Preeclampsia with nephrotic syndrome. *Kidney Int.* 13:166–77
57. Nadal, M. A., Iotti, R. M., Monserrat, A. J., Gotlieb, D., Baldi, E. M. 1971. The nephrotic syndrome of preeclampsia. *Medicina (Buenos Aires)* 39:65–75 (In Spanish)
58. See Ref. 11, pp. 77–89
59. See Ref. 11, pp. 43–76
60. See Ref. 10, pp. 199–228
61. Friedman, E. A., Neff, R. K., 1978. Hypertension-hypotension in pregnancy. Correlation with fetal outcome. *J. Am. Med. Assoc.* 239:2249–51
62. Al-Khader, A. A., Stewart, C., McGale, E. H. F., Aber, G. M. 1980. Exaggerated natriuresis in normotensive subjects with a previous history of pregnancy-associated hypertension. *Nephron* 26:23–27
63. Smeaton, J. C., Anderson, G. J., Fulton, I. S. 1977. Study of aldosterone levels in plasma during pregnancy. *J. Clin. Endocrinol. Metab.* 44:1–7
64. Nolten, W. E., Ehrlich, E. N. 1980. Sodium and mineralocorticoids in normal pregnancy. *Kidney Int.* 18:162–72
65. Benedetti, T. J., Cotton, D. B., Read, J. C., Miller, F. C. 1980. Hemodynamic observations in severe preeclampsia with a flow-directed catheter. *Am. J. Obstet. Gynecol.* 136:465–70
66. See Ref. 10, pp. 338–40
67. Gallery, E. D. M., Hunyor, S. M., Györy, A. Z. 1979. Plasma volume contraction: a significant factor in both pregnancy-associated hypertension (preeclampsia) and chronic hypertension in pregnancy. *Q. J. Med.* 192:593–602
68. Joyce, T. H. III, Debnath, K. S., Baker, E. A. 1979. Preeclampsia—relationship

of CVP and epidural analgesia. *Anesthesiology* 51:S297 (Abstr.)

69. See Ref. 10, pp. 88–118
70. Redman, C. W. G., Bonnar, J., Beilin, L. 1978. Early platelet consumption in preeclampsia. *Br. Med. J.* 1:467–69
71. Al-Khader, A. A., Stewart, C., Aber, G. M. 1979. Pregnancy-associated disease of the microcirculation. *Clin. Sci.* 56:189–92
72. Bonnar, J., Redman, C. W. G., Denson, K. W. 1976. The role of coagulation and fibrinolysis in preeclampsia. *Perspect. Nephrol. Hypertens.* 5:85–92
73. McKillop, C. A., Howie, P. W., Forbes, C. D., Prentice, C. R. M. 1977. Sequential studies in preeclampsia using plasma fibrinogen chromatography. *Br. J. Obstet. Gynaecol.* 84:909–18
74. Boneu, R., Bierme, R., Fournie, A., Pontonnier, G. 1977. Factor-VIII complex fetal growth retardation and toxemia. *Lancet* 1:263
75. Rákóczi, F., Tallián, S., Bagdány, S., Gáti, I. 1979. Platelet life span in normal pregnancy and preeclampsia as determined by a nonradioisotope technique. *Thromb. Res.* 15:553–56
76. Pollak, V. E. 1977. The role of intravascular coagulation in toxemia of pregnancy. *Perspect. Nephrol. Hypertens.* 9:107–24
77. Pritchard, J. A., Cunningham, F. G., Masson, R. A. 1976. Does coagulation have a causative role in eclampsia. *Perspect. Nephrol. Hypertens.* 5:95–101
78. Fairley, K. F., Adey, F. D., Ross, I. C., Kincaid-Smith, P. 1976. Heparin treatment in severe preeclampsia and glomerulonephritis in pregnancy. *Perspect. Nephrol. Hypertens.* 5:103–11
79. Goodlin, R. C., Haesslein, H. O., Flemming, J. 1978. Aspirin for treatment of recurrent toxemia. *Lancet* 2:51
80. Sheehan, H. L., Lynch, J. B. 1973. *Pathology of Toxaemia of Pregnancy,* pp. 524–84. Baltimore: Williams & Wilkins. 807 pp.
81. See Ref. 80, pp. 524–84
82. McKay, D. G. 1976. Disseminated intravascular coagulation and experimental disease models in pregnancy. *Perspect. Nephrol. Hypertens.* 5:401–9
83. See Ref. 10, pp. 364–67
84. Fastenberg, D. M., Fetkenhour, C. C., Choromakos, E., Shoch, D. E. 1980. Choroidal vascular changes in toxemia of pregnancy. *Am. J. Ophtalmol.* 89:362–68
85. See Ref. 45, pp. 61–106
86. Clewel, W., Meschia, G. 1976. Relationship of the metabolic clearance rate of dehydroisoandrosterone sulfate to placental blood flow. *Am. J. Obstet. Gynecol.* 125:507–8
87. Robertson, W. B., Brosens, I., Dixon, G. 1976. Maternal uterine vascular lesions in the hypertensive complications of pregnancy. *Perspect. Nephrol. Hypertens.* 5:115–26
88. Brinkman, C. R. III, Assali, N. S. 1976. Uteroplacental hemodynamic response to antihypertensive drugs in hypertensive pregnant sheep. *Perspect. Nephrol. Hypertens.* 5:363–73
89. Venuto, R., Cox, J. W., Stein, J. H., Ferris, T. F. 1976. Effect of changes in perfusion pressure on uterine blood flow in the pregnant rabbit. *J. Clin. Invest.* 57:938–44
90. Ferris, T. F. 1980. Hypertension in pregnancy. In *Hypertension Update: Mechanisms, Epidemiology, Evaluation, Management,* ed. J. C. Hunt, pp. 85–92. Summit, New Jersey Health Learning Systems. 280 pp.
91. Neuman, J., Weiss, B., Rubello, V., Cabal, L., Freeman, R. K. 1979. Diazoxide for the acute control of severe hypertension complicating pregnancy: A pilot study. *Obstet. Gynecol.* 53:505–55
92. Vink, G. J., Moodley, J., Philpot, R. H. 1980. Effect of dihydralazine on the fetus in the treatment of maternal hypertension. *Obstet. Gynecol.* 55: 519–21
93. Abitbol, M. M., Gallo, G. R., Pirani, C. L., Ober, W. B. 1976. Production of experimental toxemia in the pregnant rabbit. *Am. J. Obstet. Gynecol.* 124: 460–70
94. Abitbol, M. M., Pirani, C. L., Ober, W. B., Driscoll, S. G., Cohen, M. W. 1976. Reproduction of experimental toxemia in the pregnant dog. *Obstet. Gynecol.* 48:537–48
95. Abitbol, M. M., Ober, W. B., Gallo, G. R., Driscoll, S. G., Pirani, C. L. 1977. Experimental toxemia of pregnancy in the monkey: with a preliminary report on the renin-angiotensin system. *Am. J. Pathol.* 86:573–90
96. Jenkins, D. M., Need, J. A., Scott, J. S., Morris, H., Pepper, M. 1978. Human leukocyte antigens and mixed hymphocyte reaction in severe preeclampsia. *Br. Med. J.* 1:542–44
97. Birkeland, S. A., Kristofferson, K. 1979. Preeclampsia—a state of mother-fetus immune imbalance. *Lancet* 2: 720–23
98. Griffin, J. F. T., Wilson, E. W. 1979. Preeclampsia: a state of mother-fetus immune imbalance. *Lancet* 2:1366–67

99. Need, J. A. 1979. Immunological phenomena in preeclamptic toxemia. *Clin. Obstet. Gynaecol.* 6:443–60
100. Editorial. 1980. Genetic control of preeclampsia. *Lancet* 1:634–35
101. Beer, A. E. 1978. Possible immunological basis of preeclampsia/eclampsia. *Semin. Perinatal.* 2:39–59
102. Chesley, L. C., Annitto, J. E., Cosgrove, R. A. 1968. The familial factor in toxemia of pregnancy. *Obstet. Gynecol.* 32:303–11
103. Cooper, D. W., Liston, W. A. 1979. Genetic control of severe preeclampsia. *J. Med. Genet.* 16:409–16
104. Chesley, L. C. 1980. Hypertension in pregnancy: definitions, familial factor and remote prognoses. *Kidney Int.* 18:234–40
105. Redman, C. W. G., Bodner, J. G., Bodner, W. F., Beilin, L. J., Bonnar, J. 1978. Pre-eclampsia/antigens in severe preeclampsia. *Lancet* 2:397–99

Ann. Rev. Med. 1981. 32:291–312
Copyright © 1981 by Annual Reviews Inc. All rights reserved

PLUMONARY EDEMA DUE TO INCREASED MICROVASCULAR PERMEABILITY[1]

❖7399

Norman C. Staub, M.D.

Cardiovascular Research Institute and Department of Physiology, University of California, San Francisco, California 94143

Introduction

The only prior mention of increased lung microvascular permeability in this series is a brief description by Hopewell & Murray in their chapter, *Adult Respiratory Distress Syndrome* (72). But the topic is in vogue and there have been several recent reviews (20, 135, 145, 149, 151), as well as a minisymposium from UCLA (140). The earlier general reviews (126, 147) cover the older literature. Hurley (74) wrote an outstanding review of the subject from the experimental pathologist's point of view.

There are two reasons for the surge of interest. First, there is the increased occurrence of acute lung microvascular injury with associated pulmonary edema in many different clinical conditions (135). The morbidity and mortality from these conditions are high (72). Second, there are now specific models of acute lung microvascular injury that can yield new understanding of the early stages of increased permeability.

One of the major advances of the last decade is the development of specific, reproducible experimental models of various types of increased permeability edema. These include hemorrhagic shock (40), septicemia (21, 22), microembolism (52, 94 111, 129), neurogenic edema (16, 164), hydrochloric acid aspiration (158), drug-related edema (14, 102), and oxygen toxicity (18).

The clinical syndrome of increased permeability edema is usually characteristic. It begins with a specific episode, but the nature of the insults may

---

[1] This review and work from our laboratory were supported in part by US Public Health Service Grants HL06285 (Program Project), HL25816 (Program Project), HL14201 (Plumonary SCOR), and HL19155 (Pulmonary Vascular SCOR).

0066-4219/81/0401-0291$01.00

be diverse, such as drug intake (81), acid aspiration (23), or venous air embolism during neurosurgery (114). When edema fluid can be aspirated from deep within the lung, the protein concentration exceeds 60% of that in plasma obtained simultaneously (24, 50).

Many methods are being developed for detecting early increases in lung water content in pulmonary edema, although these are not specific for increased permeability edema. All such methods depend upon introducing a substance that in some way is diluted, concentrated, or excluded from the lung when its water content changes (153). These methods include the chest x ray (119, 146), the double indicator dilution technique (26, 86, 119), soluble gas measurement of lung tissue volume (56, 127), transthoracic electrical impedance (49), and nuclear imaging techniques (29, 66, 142, 161).

Still in their infancy are more specific methods aimed at detecting increased lung microvascular (60, 65, 122) and alveolar (77, 125) solute permeability. These new approaches may theoretically detect barrier injury before measurable amounts of fluid accumulate (147). Some of the tests have progressed from animal experimentation to preliminary human studies (61, 78).

Specific Lung Injury

Whether or not the Adult Respiratory Distress Syndrome (ARDS) constitutes a specific entity (51, 116), most pulmonary and critical care physicians agree that following severe trauma, hemorrhagic shock, sepsis, gastric fluid aspiration, viral pneumonia, narcotic abuse, etc, a certain number of patients develop progressive respiratory distress including shallow rapid breathing, dyspnea, cough, and rales. The chest x ray shows diffuse coalescing pulmonary infiltrates and these patients develop severe systemic arterial hypoxemia. The lungs become increasingly difficult to ventilate (low compliance), pulmonary vascular resistance rises, and the level of inspired oxygen must be steadily increased in order to maintain oxygen transport (4, 12, 129).

Pathologic changes in the lungs are very similar among these patients even though the inciting events may be disparate (5, 69, 124). In addition to interstitial and alveolar edema with hyaline membranes, there is widespread microembolism, infiltration of neutrophils, and destruction of type I alveolar epithelial cells. If survival is prolonged beyond several days, there is proliferation of the type II alveolar epithelial cells and, sometimes, accelerated fibrotic changes. Because the end-stage is the same, it is often assumed that a common mechanism must underlie the injury (116). But this may merely represent the stereotypic reaction of the lung to severe alveolar injury (5).

The remainder of this review deals with our current understanding of the pathogenesis of specific types of increased permeability edema.

Hemorrhagic Shock

A period of systemic arterial hypotension with mean blood pressure at or below 50 mm Hg for 1–3 h is characterized both clinically and experimentally as hemorrhagic shock. Moss (105) reviewed this topic several years ago in this series. My remarks constitute an update of his discussion and provide a starting point for the next sections. It is now clear that hemorrhagic shock by itself does not produce clinically or experimentally significant increased lung microvascular permeability, except in the dog (38). Microvascular injury with edema, when it does occur, is either the result of accompanying insults (such as soft tissue trauma or sepsis) or of overzealous fluid replacement (2, 37, 71).

There is something special about the response of the dog to hemorrhagic shock (107). Webb and his associates (166) showed by direct microcinematography of the lung surface in the anesthetized dog that changes occurred during the period of profound hypotension. These included microcirculatory stasis (probably due to intense vasoconstriction), lodgment of amorphous microemboli, focal hemorrhage, and edema. Likewise, Todd and his associates (157) reported that right duct lymph flow (believed to represent mainly lung fluid filtration) increased in anesthetized dogs during the period of hypotension, even though pulmonary vascular pressure had decreased. The same lab had previously reported, using freeze-fracture electron microscopy, that the intercellular junctions of the alveolar epithelium and to a lesser extent those of the endothelium were sufficiently altered during prolonged hemorrhagic shock to account for increased microvascular permeability (7). Recently, Michel & Hogg (101) quantified changes in the dog lung following hemorrhagic shock and resuscitation. The only significant finding, by light microscopy, was a mild leucocytic infiltration. Northrup & Humphrey (109) also measured right duct lymph flow in dogs during hemorrhagic shock and after resuscitation. Half of the dogs were "shock sensitive," meaning there were large increases in the flow of lymph, which had a high protein concentration indicative of increased microvascular permeability. Interestingly, none of their dogs showed any significant increase in extravascular lung water.

Demling and associates (42) measured lung lymph flow, pulmonary hemodynamics, and tracer protein equilibration between plasma and lung lymph in anesthetized sheep during hemorrhagic shock. They were not able to detect any significant changes in either fluid filtration or protein leakage. Demling et al (40) also studied hemorrhagic shock followed by resuscitation with the shed blood and lactated Ringer's solution in unanesthetized sheep.

After an initial decrease during the bleeding phase, lung lymph flow recovered to baseline levels by the end of the two-hour shock period. Following resuscitation lymph flow doubled but its protein concentration relative to plasma decreased. They concluded that the increased fluid filtration was due to increased microvascular pressure rather than an increase in fluid and protein permeability. A similar conclusion was reached by Malik et al (95). Even in untreated hemorrhagic shock, which led to death over 4–5 h, sheep showed little evidence of increased microvascular permeability, although lymph flow did rise terminally (39).

What is the difference between the dog and other species (sheep, humans, primates) in their response to hemorrhagic shock? Niehaus et al (108) believe it is related to reticuloendothelial system function. The sheep and man have a less active reticuloendothelial system and low circulating levels of fibronectin (opsonin) compared to the dog. In a recent comparative study Bredenberg & Webb (17) observed the responses of baboons and dogs to hemorrhagic shock. Their conclusion sums up current opinion nicely: "We suggest that 'shock lung' be considered a canine disease."

Septic Lung Injury

Of all the factors complicating shock and trauma, septicemia is the most widely accepted as a cause of increased permeability edema (1, 21, 22, 27, 38, 57, 79). One reason why septicemia is so dreaded in burned individuals is that they tend to be fluid overloaded anyway. They tolerate the excess fluid without developing overt clinical pulmonary edema unless septicemia occurs (159). What may follow is catastrophic pulmonary edema due to increased vascular permeability.

Gabel et al (58) could not detect a change in the endothelial fluid conductance during 3 h of a continuous infusion of *Escherichia coli* endotoxin in dogs. Rice et al (123) could not find evidence of pulmonary edema 4 or 24 h after *E. coli* septicemia in baboons. Sturm et al (154) did not detect increased protein permeability within 2 h of *E. coli* bacteremia in sheep.

Much of the apparent disagreement about septic lung injury seems to be due to the time after the initial insult at which measurements were made, the amount of endotoxin or bacteria given, and whether one is monitoring pulmonary hemodynamics or fluid and protein exchange. The development of the sheep model of septic lung injury—the *Pseudomonas* bacteremia model (21) and, more recently, the *E. coli* endotoxemia model (22, 41)—permits some clarification of these factors.

There are two fairly distinct phases to the pulmonary vascular effect of septicemia. First is a period of pulmonary hypertension that occurs during or shortly after the insult and lasts about 2 h. If the septic insult is massive, the first phase may be the only one seen (47). The initial phase is characterized by an increased lung lymph flow and a decrease in the lung lymph protein concentration (21, 22, 38). Although endothelial barrier permeability may be increased, the predominant effect appears to be hemodynamic. That is, increases in fluid filtration are due mainly to increased hydrostatic pressure in those portions of the pulmonary microcirculation being perfused (154).

Some of the more seriously injured regions of lung may not be perfused because of intense vasoconstriction and physical obstruction by bacterial clumps. Demling & Flynn (38) showed an increase in the erythrocyte concentration in lung lymph during this phase, which is often indicative of increased microvascular pressure. Severe respiratory insufficiency may arise from terminal airway constriction and marked ventilation/perfusion mismatching, apparently caused by the same agent that promotes the vasoconstriction. Interestingly, the first phase can be almost completely blocked by inhibitors of prostaglandin synthesis (38, 110).

The second phase develops 2–5 h after the initial injury. We first reported the delayed effect in the *Pseudomonas* model (21), but it is also seen with the more refined endotoxin model (22, 38, 70). This phase is characterized by a further increase in lung lymph flow (sometimes to very high levels) at a time when the hemodynamic effects (namely, the pulmonary hypertension and obstruction) are subsiding. The high lymph flow is accompanied by an increase in the lymph protein concentration indicative of increased permeability of the endothelial barrier. There are two reasons for the delayed change: First, as the vasoconstriction and obstruction abate the vascular surface area for exchange increases. Second, it appears that some metabolic process must occur before the full increased permeability injury is manifest. Polymorphonuclear leucocytes appear to play a central role in the development of septicemic injury (67, 84), although leucocytes have little to do with the hypertensive phase (53, 118).

In experimental animals, of course, the injury is deliberately kept minimal, is completely reversible, and most of the animals survive (21, 22, 41). Sometimes, however, it is fatal mainly because of severe respiratory failure and systemic collapse. Indeed, two of our original seven *Pseudomonas* sheep died this way and, recently, Esbenshade et al (47) reported the fatal effects of large doses of endotoxin in sheep.

The mechanism of the hypertensive phase includes some physical obstruction but, more importantly, intense pulmonary artery vasoconstriction mediated by one or more of the vasoconstrictor prostaglandins. Thromboxane B_2, the metabolite of the potent vasoconstrictor thromboxane A_2, is increased in plasma and lung lymph and correlates closely with the severity of the pulmonary hypertension (38). Whether the source of the thromboxane is platelets or endothelial cells is not clear. The vasoconstrictor response can be blocked by cyclo-oxygenase inhibitors without interfering with the increased permeability phase (110).

Since the response occurs slowly, the mechanism of the second phase involves some metabolic interaction that also develops slowly (21). There is good evidence that the polymorphonuclear leucocytes (PMNs) are involved since the second-phase effects are largely prevented in neutropenic animals (67). Since PMNs are also implicated in lung injury following embolism (52), dialysis (32), and oxygen toxicity (54) (see following sections), they are of more than passing interest. The exact nature of the injury induced by the PMNs is not entirely clear and has been variously attributed to physical migration through the vascular endothelium (68), to release of lysosomal enzymes (41), or to superoxide radicals (128). (For review, see 151). Pharmacologic doses of corticosteroids (methylprednisolone, 30 mg/kg) have a protective effect against *E. coli* endotoxin injury in sheep (15) and against septicemia in man (139). This is consistent with the delayed nature of the increased permeability phase. Neither in man (139) nor in sheep (15) do corticosteroids have any significant effect after the increased permeability is fully developed.

Microembolism Syndrome

Since evidence of widespread microemboli is found so frequently in the lungs of people dying of ARDS (104, 106, 167), and since clinical evidence of rising pulmonary vascular resistance has such serious import in critically ill patients (12), there has been much investigation into the role of various types of microemboli in the pathogenesis of increased permeability edema (111, 129). Even if microemboli turn out to be of secondary importance, it is essential to know the mechanism of the injury that they may produce in their own right.

We are especially interested in the air embolism model of increased permeability edema in sheep (52). It also has a clinical counterpart. Venous air emboli during decompression are very common and occasionally cause acute fatal pulmonary edema, as well as a syndrome known as the "chokes" (44). Air emboli also occur during neurosurgery and may promote acute pulmonary edema (75, 114).

Not everyone agrees that all microemboli cause serious lung injury. McKeen et al (99) showed that intravenous infusion of 10% fat emulsion in sheep caused physical obstruction but little evidence of injury to the microvascular barrier. Barie et al, however, disagree (6). The study of pathogenesis of the lung injury associated with microemboli revolves around four major possibilities, namely, intravascular coagulation (129), platelets (163), physical injury (111), and PMNs (52).

The essential role of intravascular coagulation in lung injury following trauma, sepsis, and shock is espoused by Saldeen and his group in Sweden (87, 103, 129, 130). Malik and associates (93, 94) support the importance of fibrin and its degradation products. The evidence is certainly impressive. Saldeen early on made the clinical association between the amount of fibrin in pulmonary vessels and the development of ARDS or, as he prefers to call it, the microembolism syndrome. Heparin to prevent coagulation, proteolytic enzymes to increase plasmin and hasten fibinolysis, and depletion of fibrinogen in experimental animals largely prevented the microembolism syndrome. These results are to be expected since in Saldeen's experimental work embolization is induced by intravenous thrombin. But Malik & van der Zee (94) showed that heparin and fibrinogen depletion protect dogs following glass microemboli. Unfortunately, Binder et al (10) could not demonstrate that fibrinogen depletion or heparin protected sheep.

Saldeen (130) believes the real culprit in the microembolism syndrome is a specific inhibitor of fibrinolysis. In this view, during fibrinolysis fibrin degradation products are formed which, if given the opportunity, increase microvascular permeability. Inhibition of fibrinolysis promotes injury because the slowed rate allows a prolonged release of fibrinolytic peptides thereby giving time for the endothelial injury to develop. In other words, too rapid fibrinolysis does not cause increased permeability; neither does no fibrinolysis at all.

Vaage and his associates (162, 163) showed that platelet embolization in the lung will promote an increase in vascular leakiness. But the platelet effect is transient and does not lead to prolonged permeability changes (151). These results have been confirmed (11, 53) and are very important in the overall picture of acute lung injury. There is little doubt that part of the initial injury in several experimental models is due to platelet embolization of the lung.

The heart-lung bypass literature reveals that blood flowing through membrane oxygenators and returning to a patient or experimental animal causes an immediate pulmonary hypertension and increased microvascular permeability (113). This has been well-documented by Delaney et al (36) in unanesthetized sheep. They showed that immediately after recirculation of blood through an extracorporeal device there was an increase in lung lymph and protein flow that lasted for two or more hours. Following this, although blood continued to flow through the oxygenator, there was no further effect. Binder et al (11) found that platelet embolization of the lung, using antiplatelet serum, caused an immediate marked pulmonary hypertension and a brief period of increased fluid and protein exchange in sheep. Once the platelets had been depleted, however, embolization still caused the delayed increase in microvascular permeability. This description matches that of septic lung injury. The pulmonary response to microemboli also has an early hypertensive phase and a delayed increased permeability phase. Short-term experiments on embolization (94) lead to somewhat different conclusions than long-term experiments (111).

Fountain et al (53) compared the role of platelets and leucocytes in the hypertensive phase of extracorporeal oxygenator-induced injury. They found that although there was a profound leucopenia with marked retention of PMNs in the lung, sulfinpyrazone, an antiplatelet aggregation agent, blocked almost all of the plumonary hypertension without affecting the leucocyte sequestration. Several other workers have shown beneficial effects of agents that interfere with platelet aggregation (90, 120). The pulmonary microvascular response to several agents depends upon a time-related interaction of various elements.

When we found we could not prevent the microembolism-induced permeability increase by fibrinogen or platelet depletion, we speculated that physical injury might account for the syndrome (111): high distending pressure together with a high linear velocity of blood flow through the restricted microcirculation might be sufficient to injure the endothelial barrier. New evidence, however, leads us to abandon this concept, at least in its original form. Physical restriction of the pulmonary microcirculation by successive lobar resection (85) does not cause an increased permeability edema. It causes edema, but it is high pressure edema (see the section on high altitude edema). More germane is the finding that leucocyte depletion prevents most of the permeability increase after emboli even though the obstruction and high microvascular pressure still occur (52).

A few additional words are in order about neutrophils in the lung. Normally there are large numbers of neutrophils in the capillaries and marginated in the venules in the lung (151) but few neutrophils outside the vascular compartment (30). Apparently, these do no harm. In the acute phase of ARDS, however, the lung interstitium is full of neutrophils. Several workers have related neutrophil sequestration in the lung to injury (97, 98). Recently, Johnson & Malik (76) confirmed the necessary neutrophil component of acute lung injury. Since Malik has been a strong proponent of the fibrin injury theory, a reconciliation of these two positions is needed. He suggests an interaction between intravascular clotting and neutrophils. Unfortunately, Lindquist in Saldeen's lab did not find that leucopenia affected the lung injury due to thrombin-induced fibrin microemboli (87).

Our own experiments on microemboli in anesthetized and unanesthetized (52) sheep point strongly to leucocyte involvement. Depletion of leucocytes by nitrogen mustard eliminated a large part of the increased permeability response to microemboli. Leucocyte depletion did not affect the hypertensive phase.

A word of caution is necessary. Neutrophil sequestration in the lungs, no matter how dramatic, does not necessarily cause injury (53, 102). The lung injury due to ethchlorvynol is a direct toxic injury to the endothelium, but it causes profound leucocyte retention in the lungs (48).

Many workers (36, 68) believe that microvascular injury by neutrophils requires their activation by complement. This provocative idea needs to be confirmed in each type of lung injury in which PMNs are reputed to be the active agent.

Although the microembolism syndrome is important in its own right, is it sufficient to account for the ARDS syndrome, as Saldeen believes (129)? In experimental animals, dynamic lung compliance is decreased and the alveolar-arterial oxygen tension difference is increased. There is an increase in pulmonary vascular resistance and an increase in fluid and protein permeability. But these tend to be limited to several hours duration and infrequently result in death, suggesting that additional factors remain to be discovered.

Hydrochloric Acid Aspiration

Aspiration of acid or gastric contents (Mendelson's syndrome) has marked and immediate pulmonary effects (23). Although the syndrome has been known for over 30 years and intensive modern therapy has been applied, the mortality rate is still substantial. My discussion is limited to the effects of hydrogen ion on alveolar-capillary permeability. The acid injury is virtually instantaneous because lung fluids rapidly buffer the H^+ to a harmless level ($pH > 2.5$) (77). No injury is caused if the aspirate $pH > 2.0$ (0.01 N).

Several experimental studies of increased permeability injury caused by 0.1 N HCl have been published recently, and the data about its early natural course are in agreement. A rather large inoculum ($\geqslant 2$ ml/kg) is necessary to cause significant injury. A large volume of 0.1 N HCl appears to be more injurious than a small volume of 1 N HCl. This may be because the more concentrated acid injures a much smaller area of lung and literally cauterizes the tissues thereby sealing off the destroyed area.

The acid injury is immediate and direct (74). Edema begins to form rapidly as shown by continuous measurements of weight gain in lung lobes (158), gross edema developing over 2–3 h (107), and a rapid increase in the clearance of low molecular weight tracers (such as ^{99}Tc-DTPA) from fluid instilled into the air spaces (77).

Substantial controversy revolves around whether there is any specific treatment, particularly whether concentrated albumin solutions or high dose corticosteroids are useful. Toung et al (158), using the isolated perfused dog lung lobe, found that large amounts of concentrated human serum albumin given within 3–5 min of 1 N HCl aspiration were nearly 100% effective in preventing the increased permeability edema. Neither Grimbert and associates (62) nor Nanjo in our laboratory (107) were able to find any beneficial effect of concentrated albumin solutions given intravenously, either before or after 0.1 N HCl aspiration, to anesthetized dogs with an open thorax, self-perfused, weighed lung lobe. Thus, in intact animals, there is no evidence that treatment with concentrated albumin solutions is any more useful than in other forms of increased permeability edema.

Toung et al found that high dose corticosteroids were beneficial in preventing the increased permeability injury, if they were given shortly after the acid aspiration (< 5 min). This use of corticosteroids has been examined both clinically and experimentally. Clinical studies have not detected any beneficial effect (23), but these studies are always complicated by other features of the injury and by the difficulty of equating individual case reports. The animal studies have also been controversial. Glauser et al (59) failed to show any beneficial effects of corticosteroids following acid injury in dogs. His group measured alveolar-capillary membrane permeability to small inert tracer molecules placed in the air spaces. Shellito & Murphy (137) found that corticosteroids do have beneficial effects after acid aspiration: methylprednisolone prevented PMN infiltration into the acid-injured lobe and favorably altered alveolar macrophage function. Although PMNs are present in large numbers during the late inflammatory reaction to acid, they are not involved in the immediate chemical injury. This is conclusively proven by the completely isolated, perfused lung lobe experiments in which there are no circulating PMNs.

Neurogenic Pulmonary Edema

Instantaneous pulmonary edema due to brain trauma (141) is probably a myth (147), but acute pulmonary edema does occur, albeit not commonly, in association with severe head injury, subarachnoid hemorrhage, increased cerebrospinal fluid pressure, and convulsions. It is called neurogenic because of the apparent dominant role played by α-adrenergic discharge. Neurogenic edema was believed (8, 89, 132) to be of the high pressure type and due principally to the massive α-adrenergic discharge, which caused intense peripheral vasoconstriction, displacement of blood to the central circulation, and changes in left ventricular wall stiffness that increased left atrial pressure. These events and the associated edema were largely prevented by α-blockade.

In a provocative paper, Theodore & Robin (155) reviewed the clinical and experimental data and suggested that neurogenic pulmonary edema is probably of the increased microvascular permeability type. They speculated that the endothelial barrier was injured by transient overpressurization. There was little evidence of increased pulmonary vascular pressure in the clinical literature on neurogenic edema. They reasoned that this is exactly what one would expect if the massive peripheral vasoconstrictor effects caused only a transient period of very high pulmonary vascular pressure.

Their theory fits the concept of pressure-induced permeability increases (pore-stretching) proposed by Shirley et al (138) and supported by the data of Schneeberger & Karnovsky (133) and of Pietra et al (117). Although we have never found any evidence of pore-stretching in pulmonary edema due to high pressure (45, 85, 100), that could be because we never raised the pressure high enough in our steady-state experiments. In transient conditions much higher pressures could be obtained.

Wray & Nicotra (170) reported one patient in whom continuous pressure measurements showed transient episodes of high pulmonary arterial and wedge pressure, apparently supporting Theodore & Robin. Carlson et al (25) reported one patient with acute pulmonary edema following intracranial hemorrhage in whom very high pulmonary vascular pressures were also reported. They found, in samples of edema fluid, that plasma protein concentration ratio average 0.76; they argued that the edema in their patient could be explained on the basis of the high pressure rather than increased permeability. Their conclusion surprises me since the high protein concentration in the edema fluid (up to 0.85 of the plasma level at one point) indicated increased endothelial barrier permeability, especially when compared to the large differences measured in edema fluid protein concentrations in patients with hemodynamic and increased permeability edemas by the same workers (24).

Neither of the patient reports can be used to confirm or deny a particular theory of the pathogenesis of neurogenic pulmonary edema. Indeed, high pulmonary vascular pressure may be a very important component of the process. According to Starling's equation, fluid filtration under any circumstances is proportional to and driven mainly by the microvascular hydrostatic pressure (148).

New experimental data, however, does not entirely fit the high pressure theory. Two groups (16, 92) raised cerebrospinal fluid (CSF) pressure in anesthetized sheep; both recorded pulmonary arterial and left atrial pressures continuously. Although lung lymph and protein flow were moderately increased, they did not record any high vascular pressures, transient or otherwise. One can argue that the general anesthesia and trauma of surgery (both necessary for these experiments) blunted the sympathetic discharge. But one cannot explain away the fact that lymph and protein flow increased without any significant pressure changes. The authors of these papers intimate that there may be a direct effect of sympathetic discharge within the pulmonary circulation. Malik's group (64) followed this idea further by stimulating the sympathetic efferent nerves to the lungs (and other intrathoracic organs) directly in anesthetized sheep. Lung lymph and protein flow increased slightly, but pulmonary artery pressure and cardiac output also increased. Thus, they could not clearly separate hemodynamic from permeability changes.

Another very interesting form of neurogenic edema is that associated with convulsions (9, 88, 131), which can also be prevented by α-blockade. This phenomenon was reproduced in the anesthetized, paralyzed sheep by using the convulsant bicuculline (143). In the only data thus far reported, transient large increases in pulmonary vascular pressure usually occurred during convulsions and seemed to be correlated with the subsequent increases in lung lymph and protein flow. These data, if confirmed, clearly support the hypothesis of Theodore & Robin.

Although α-blockade is the usual means of blocking neurogenic edema, naloxone appears to be partially effective in preventing the lymph and protein flow increases seen in sheep during elevation of CSF pressure (115).

In summary, neurogenic pulmonary edema appears to be of the increased permeability variety, but its genesis is more complex than has been supposed. High vascular pressure appears to be coupled to some intrapulmonary effect of sympathetic outflow on endothelial barrier conductance.

Oxygen Toxicity

The dilemma of oxygen need and oxygen toxicity in patients with acutely injured lungs is real enough. There can be no doubt that the prolonged breathing of oxygen-rich mixtures harms the lungs of experimental animals. Presumably this holds true for man as well. Current practice is to limit inspired oxygen to the lowest fraction and the shortest duration necessary to obtain adequate systemic oxygen delivery. Exactly what injury is caused by oxygen, when it occurs, and how it can be prevented or treated is the subject of much investigation (28, 33, 156, 169).

Normal experimental animals breathing 100% oxygen show a fairly stereotypic sequence of biochemical, histological, and functional changes terminating in death, usually within 3–5 days (18, 82). Monkeys may be an exception; although they get very sick and show interstitial edema, half of them survive and recover reasonably normal lung function (79). In all species studied, however, the primary intrapulmonary site of injury appears to be the microvascular endothelium.

Humans also show evidence of pulmonary abnormalities during exposure to high oxygen mixtures (28, 31). Although such exposure is not allowed to continue for long periods, the outcome would likely be similar to that of other animals. The question is whether humans are more like monkeys (79) or rats (35).

The study of oxygen toxicity in association with acute lung injury is complicated. Some pathologists feel there is evidence for insult added to injury (121). Tierney et al (156) reviewed some of the interactions that affect the lung's sensitivity to oxygen. Some of these increase sensitivity (paraquat and influenza pneumonia) but others decrease sensitivity (diet, prior exposure to 0.85 atm oxygen). Attempting to sort out the contribution of high inspired oxygen tension to the pathologic picture of acute lung injury is probably an exercise in futility (5, 69).

The biochemical lesion induced by oxygen appears to be related to the overwhelming of the body's defenses against superoxide radical formation (33, 169). Briefly, in the course of normal oxidative metabolism, some highly active oxygen radicals (superoxide, hydrogen peroxide, hydroxyl ion) are produced but these are scavenged by the ubiquitous intracellular enzyme, superoxide dismutase. The rate of oxygen free-radical formation is increased by raising the partial pressure of oxygen, although it is not clear whether that is sufficient to saturate the dismutase pathway. Why is there a relatively long interval between the onset of oxygen breathing and the final toxic injury? Why can animals survive in 85% oxygen, but die breathing 100% oxygen (35)? Perhaps additional factors come into play that ultimately spell defeat for the cellular defense mechanisms.

Attempts to determine very early biochemical changes in the lung as a result of oxygen breathing are inconclusive. Likewise, attempts to detect early ultrastructural changes are equivocal. One group has obtained evidence for a modest increase in alveolar barrier solute permeability at about 60 h (96).

Important new data in lung fluid balance and microvascular permeability have been obtained. Lambs 2–4 weeks old were exposed to 100% oxygen for several days (18). Aside from a modest fall in pulmonary artery pressure and cardiac output at the onset of the exposure, no significant changes in hemodynamics were observed. There were also no changes in lung lymph flow, lymph protein concentration, or animal behavior until the third or fourth day, at which time lymph flow began to increase as did its protein concentration. Shortly thereafter the animals became sick and ultimately died. Mysteriously, the main terminal gas exchange problem was an acute respiratory insufficiency as indicated by a rising arterial CO_2 tension. There was almost no stimulation to breathing. Arterial oxygen tension was well maintained until near death. The amount of lung edema, estimated histologically and by the measurement of extravascular lung water content, seemed insufficient to cause death. Again, this suggests some additional effect of high oxygen concentration. The sequence of events in oxygen toxicity in adult sheep is very similar to that in the newborn lamb (46).

Fox et al (54) found that in rabbits breathing 100% oxygen there is a terminal influx of PMNs on the third day. Shasby et al (136) reported that the mortality rate from oxygen breathing was reduced by half when rabbits were made neutropenic before oxygen exposure. In those animals that died, the lung showed very little evidence of edema or oxygen injury. The PMNs obviously had something to do with the lung injury, just as they do with endotoxemia and microemboli. But we cannot get away from the fact that other factors must be important since a third of the rabbits died anyway.

Tolerance means the development of resistance to oxygen breathing by some preliminary maneuver (33). Tolerance can be induced by prolonged breathing of 85% oxygen and by prior exposure to other agents, such as nitrogen dioxide (34) and endotoxin (55). The most noticeable biochemical event in acquiring tolerance is an increase in the tissue concentration of the superoxide dismutase (33, 169). The association of increased levels of this enzyme with protection, and the possible role of PMNs in producing superoxide radicals (128), is obvious. But whether there is a cause and effect relationship remains to be proven. Certainly, the concept of a terminal influx of neutrophils coupled with an increase in superoxide radical formation is a tantalizing explanation for the sudden turn of events that occurs on about the third day.

The acute lung injury, edema, and death following inhalation of hyperbaric oxygen (i.e. pressure $\geqslant 1$ atm) is quite a different story (9). This form of oxygen toxicity occurs quickly and is unlikely to be due to direct lung injury. It appears to result chiefly from the central nervous system effects of high oxygen pressure. Convulsions occur and the lung injury may represent a form of neurogenic pulmonary edema (9, 28).

High Altitude Pulmonary Edema

The mechanisms of high altitude pulmonary edema and of pulmonary vasoconstriction in response to acute alveolar hypoxia are unknown, but one thing is certain. The former depends upon the latter; not necessarily as cause and effect but as a necessary antecedent.

Mountain climbers, physicians, and investigators continue to be fascinated by this illness, a fascination out of all proportion to its occurrence in man. The incidence in adults over 21 years of age is very low (73). It is more common in the young (91, 134) although the explanation for this is not clear. Perhaps it is because the young are more active or have an excessive vasoconstrictor response to hypoxia. Those who have had an episode of high altitude edema are very susceptible to its recurrence; thus something about the individual's response to hypoxia is the determining factor. High altitude edema occurs more frequently in fully acclimatized individuals who have descended to sea level and returned quickly to high altitude, but it also occurs in individuals who have never been at high altitude before (144). Coates et al (29) believe that interstitial edema occurs fairly commonly in individuals exposed to low barometric pressure. Their evidence, obtained through indirect nuclear medicine methods to assay lung density in volunteers in a hypobaric chamber, is suggestive.

High altitude pulmonary edema is classed as an increased permeability edema for two reasons. First, there is little evidence of heart failure or elevation of pulmonary wedge pressure (73). Second, in the older literature, the lungs of patients dying with high altitude edema were reported to have "proteinaceous exudate filling the alveoli" (3, 112). There is apparently no literature on the protein content of high altitude edema fluid, although there is an unpublished report (Wilson and Hultgren, personal communication) of a sample of edema fluid and blood plasma obtained from a climber on Mt. McKinley in Alaska. The protein concentration of the edema fluid was 60% of that in plasma, definitely higher than that seen in patients with high pressure edema (24, 50) but borderline for a conclusive diagnosis of increased permeability edema. The data fit into that gray zone where a combination of increased permeability and high pressure edema may coexist. This is quite reasonable, of course, because in high altitude edema the pulmonary artery pressure is very high and, presumably, the pressure is also high in some parts of the microcirculation.

The leading theory to explain high altitude edema is the "over-perfusion concept," most vigorously supported by Hultgren (73). According to this theory, the hypoxic pulmonary arterial vasoconstriction is so intense that it virtually shuts off a large portion of the pulmonary microcirculation, thereby restricting blood flow to a small fraction of the lung (165). Cardiac output is maintained and may even be augmented because a period of exercise frequently precedes the development of the edema. Thus, the entire cardiac output is forced through a restricted microvascular bed, and the high pulmonary artery pressure is transmitted to the perfused portion of the microcirculation and somehow causes the endothelial barrier injury. In this respect, high altitude pulmonary edema is like the microembolism syndrome (111, 152).

The report of four individuals with congenital absence of the right pulmonary artery who developed pulmonary edema in high altitude is provocative (63). The data certainly appear to support the over-perfusion theory, but it doesn't convincingly explain the injury (150). It takes more than high pressure and high linear blood flow velocity in the pulmonary microcirculation to damage the endothelial barrier (85). The acute alveolar hypoxia, which by itself does not alter microvascular permeability (13), may have an effect in association with high distending pressure and high shear forces.

The possible role of neutrophil activation in the lung should not be ignored. Some of the clinical and pathologic descriptions of high altitude edema refer to unexplained bronchopneumonia (3, 112, 168) and a peripheral leucocytosis (83, 91).

Literature Cited

1. Anderson, R. R., Holliday, R. L., Driedger, A. A., Lefcoe, M., Reid, B., Sibbald, W. J. 1979. Documentation of pulmonary capillary-permeability in the adult respiratory-distress syndrome accompanying human sepsis. *Am. Rev. Resp. Dis.* 119:869–77
2. Anderson, R. W., DeVries, W. C. 1976. Transvascular fluid and protein dynamics in the lung following hemorrhagic shock. *J. Surg. Res.* 20:281–90
3. Arias-Stella, J., Kruger, H. 1963. Pathology of high altitude pulmonary edema. *Arch. Pathol.* 76:147–57
4. Ashbaugh, D. G., Bigelow, D. B., Petty, T. L., Levine, B. E. 1967. Acute respiratory distress in adults. *Lancet* 2:319–23
5. Bachofen, M., Weibel, E. R. 1977. Alterations of the gas exchange apparatus in adult respiratory insufficiency associated with septicemia. *Am. Rev. Resp. Dis.* 116:589–615
6. Barie, P. S., Gertzberg, N. B., Neumann, P. H., Hilpert, E. E., Malik, A. B. 1980. Pulmonary edema and intravascular coagulation following fat microembolization in sheep. *Am. Rev. Resp. Dis.* 121(Pt. 2):424
7. Barrios, R., Inoue, S., Hogg, J. C. 1977. Intercellular junctions in "shock lung." *Lab. Invest.* 36:628–35
8. Bean, J. W., Beckman, D. L. 1969. Centrogenic pulmonary pathology in mechanical head injury. *J. Appl. Physiol.* 27:807–12
9. Bean, J. W., Zee, D., Thom, B. 1966. Pulmonary changes with convulsions induced by drugs and oxygen at high pressure. *J. Appl. Physiol.* 21:865–72
10. Binder, A. S., Nakahara, K., Ohkuda, K., Kageler, W., Staub, N. C. 1979. Effect of heparin and fibrinogen depletion on lung vascular permeability after emboli. *J. Appl. Physiol.* 47:213–19
11. Binder, A. S., Kageler, W., Perel, A., Flick, M. R., Staub, N. C. 1980. Effect of platelet depletion on lung vascular permeability after microemboli in sheep. *J. Appl. Physiol.* 48:414–20
12. Blaisdell, F. W., Schlobohm, R. M. 1973. The respiratory distress syndrome: A review. *Surgery* 74:251–62
13. Bland, R. D., Demling, R. H., Selinger, S. L., Staub, N. C. 1976. Effects of alveolar hypoxia on lung fluid and protein transport in unanesthetized sheep. *Circ. Res.* 40:269–74
14. Bowers, R. E., Brigham, K. L., Owen, P. J. 1977. Salicylate pulmonary edema: mechanism in sheep and review of the clinical literature. *Am. Rev. Resp. Dis.* 115:261–68
15. Bowers, R. E., Haynes, J., McKeen, C., Brigham, K. L. 1978. Methylprednisolone prevents endotoxin-induced high lung vascular permeability in awake sheep. *Clin. Res.* 26:444A
16. Bowers, R. E., McKeen, C. R., Park, B. E., Brigham, K. L. 1979. Increased pulmonary vascular-permeability follows intracranial hypertension in sheep. *Am. Rev. Resp. Dis.* 119:637–41
17. Bredenberg, C. R., Webb, W. R. 1979. The pulmonary response to hemorrhagic shock; differences between primate and dog. *Circ. Shock* 6:165
18. Bressack, M. A., McMillan, D. D., Bland, R. D. 1979. Pulmonary oxygen-toxicity: increased microvascular permeability to protein in unanesthetized lambs. *Lymphology* 12:133–39
19. Deleted in proof
20. Brigham, K. L. 1979. Pulmonary edema: cardiac and noncardiac. *Am. J. Surg.* 138:361–67
21. Brigham, K. L., Woolverton, W. C., Blake, L. H., Staub, N. C. 1974. Increased sheep lung vascular permeability caused by pseudomonas bacteremia. *J. Clin. Invest.* 54:792–804
22. Brigham, K. L., Bowers, R. E., Haynes, J. 1979. Increased sheep lung vascular

permeability caused by *Escherichia coli* endotoxin *Circ. Res.* 45:292–97

23. Bynum, L. J., Pierce, A. K. 1976. Pulmonary aspiration of gastric contents. *Am. Rev. Resp. Dis.* 114:1129–36
24. Carlson, R. W., Schaeffer, R. C., Michaels, S. G., Weil, M. H. 1979. Pulmonary-edema fluid—spectrum of features in 37 patients. *Circulation* 60:1161–69
25. Carlson, R. W., Schaeffer, R. C., Michaels, S. G., Weil, M. H. 1979. Pulmonary edema following intracranial hemorrhage. *Chest* 76:731–34
26. Chinard, F. P. 1975. Estimation of extravascular lung water by indicator-dilution techniques. *Circ. Res.* 37: 137–45
27. Christy, J. H. 1971. Pathophysiology of gram-negative shock. *Am. Heart J.* 81:694–701
28. Clark, J. M., Lambertsen, C. J. 1971. Pulmonary oxygen toxicity. *Pharmacol. Rev.* 23:37–117
29. Coates, G., Gray, G., Mansell, A., Nahmias, C., Powles, A., Sutton, J., Webber, C. 1979. Changes in lung volume, lung density and distribution of ventilation during hypobaric decompression. *J. Appl. Physiol.* 46:752–55
30. Cohen, A. B., Batra, G., Petersen, R., Podany, J., Nguyen, D. 1979. Size of the pool of alveolar neutrophils in normal rabbit lungs. *J. Appl. Physiol.* 47:440–44
31. Comroe, J. H. Jr., Dripps, R. D., Dumke, P. 1945. Oxygen toxicity, *J. Am. Med. Assoc* 128:710–17
32. Craddock, P. R., Fehr, J., Brigham, K. L., Kronenbert, R. S. Jacob, H. S. 1977. Complement and leucocyte-mediated pulmonary dysfunction in hemodialysis. *N. Engl. J. Med.* 296:769–74
33. Crapo, J. D., Peters-Golden, M., Marsh-Salin, J., Shelburne, J. S. 1978. Pathologic-changes in lungs of oxygen-adapted rats. *Lab. Invest.* 39:640–53
34. Crapo, J. D., Sjostrom, K., Drew, R. T. 1978. Tolerance and cross-tolerance using NO_2 and O_2. I. Toxicology and biochemistry. *J. Appl. Physiol.* 44:364–69
35. Crapo, J. D., Barry, B. E., Foscue, H. A., Shelburne, J. 1980. Structural and biochemical changes in rat lungs occurring during exposures to lethal and adaptive doses of oxygen. *Am. Rev. Resp. Dis.* 122:123–43
36. Delaney, A. G., Zapol, M., Erdmann, A. J. III 1979. Lung transvascular fluid dynamics with extracorporeal membrane oxygenation in unanesthetized lambs. *J. Thorac. Cardiovasc. Surg.* 77:252–58
37. Demling, R. H. 1980. Lung fluid and protein dynamics during hemorrhagic shock resuscitation and recovery. *Circ. Shock* 7:149–62
38. Demling, R. H., Flynn, J. 1980. Humoral factors and lung injury during shock, trauma and sepsis. In *Cell Injury in Shock, Anorexia and Ischemia* ed. B. Trump, R. Cowley. Baltimore: Williams & Wilkins
39. Demling, R. H., Manohar, M., Will, J. A. 1979. Effect of untreated hemorrhagic-shock on lung fluid and protein dynamics in the unanesthetized sheep. *Microvasc. Res.* 17:S136
40. Demling, R. H., Niehaus, G., Will, J. A. 1979. Pulmonary microvascular response to hemorrhagic shock, resuscitation, and recovery. *J. Appl. Physiol.* 46:498–503
41. Demling, R. H., Proctor, R., Day, N., Starling, J. R. 1980. Lung lysosomal enzyme release during hemorrhagic shock and endotoxemia. *J. Surg. Res.* 28:269–79
42. Demling, R. H., Selinger, S. L., Bland, R. D., Staub, N. C. 1975. Effect of acute hemorrhagic shock on pulmonary microvascular fluid filtration and protein permeability in sheep. *Surgery* 77: 512–19
43. Deleted in proof
44. Elliott, D. H., Hallenbeck, J. M. 1975. The pathophysiology of decompression sickness. In *The Physiology and Medicine of Diving,* ed. P. B. Bennett, D. H. Elliott, pp. 435–55. London: Baillieré Tindall
45. Erdmann, J. A. III, Vaughan, T. Jr., Brigham, K. L., Woolverton, W. C., Staub, N. C. 1975. Effect of increased vascular pressure on lung fluid balance in unanesthetized sheep. *Circ. Res.* 37:271–84
46. Erdmann, J. A. III, Zapol, W. M. 1980. 100% oxygen breathing increases lung capillary permeability and lung water in awake sheep. *Am. Rev. Resp. Dis.* 121(Pt. 2):430
47. Esbenshade, A. M., Newman, J. H., Lamsand, P. M., Brigham, K. L. 1980. High dose endotoxin in sheep: a model of primary pulmonary edema with respiratory failure. *Am. Rev. Resp. Dis.* 121(pt. 2):430
48. Fairman, R. P., Glauser, F. L., Falls, R. 1980. Drug induced (ethchlorvynol) increased lung vascular permeability. *J. Appl. Physiol.* In press
49. Fein, A., Grossman, R. F., Jones, J. G., Goodman, P. C., Murray, J. F. 1979. Evaluation of transthoracic electrical

impedance in the diagnosis of pulmonary edema. *Circulation* 60:1156–60

50. Fein, A., Grossman, R. F., Jones, J. G., Overland, E., Pitts, L., Murray, J. F., Staub, N. C. 1979. The value of edema fluid protein measurement in patients with pulmonary edema. *Am. J. Med.* 67:32–38
51. Fishman, A. P. 1973. Shock lung: a distinctive nonentity. *Circulation* 47: 921–23
52. Flick, M. R., Perel, A., Staub, N. C. 1980. Leucocytes are required for increased lung microvascular permeability after microemboli in sheep. *Circ. Res.* In press
53. Fountain, S. W., Martin, B. A., Musclow, E., Cooper, J. D. 1980. Pulmonary leukostasis and its relationship to pulmonary dysfunction in sheep and rabbits. *Circ. Res.* 46:175–80
54. Fox, R. B., Hoidal, J. R., Brown, D. M., Repine, J. E. 1980. Hyperoxia causes a preterminal influx of polymorphonuclear leucocytes (PMN) into the lungs and is associated with increased lung lavage chemotaxin (CTX) for PMN and death of alveolar macrophages (AM). *Am. Rev. Resp. Dis.* 121(pt. 2):340
55. Frank, L., Summerville, J. 1980. A mechanism of protection from pulmonary oxygen toxicity by endotoxin (E) treatment. *Am. Rev. Resp. Dis.* 121:340
56. Friedman, M., Kaufman, S. H., Wilkins, S. A. 1980. Analysis of rebreathing measurements of pulmonary tissue volume in pulmonary edema. *J. Appl. Physiol.* 48:66–71
57. Fulton, R. L., Jones, C. E. 1975. The cause of post traumatic pulmonary insufficiency in man. *Surg. Gynecol. Obstet.* 140:179–86
58. Gabel, J. C., Drake, R. E., Arens, J. F., Taylor A. E. 1978. Unchanged pulmonary capillary filtration coefficients after *Escherichia coli* endotoxin infusion. *J. Surg. Res.* 25:97–104
59. Glauser, F. L., Millen, J. E., Falls, R. 1979. Increased alveolar epithelial permeability with acid aspiration: The effects of high-dose steroids. *Am. Rev. Resp. Dis.* 120:1119–23
60. Gorin, A. B., Weidner, W. J., Demling, R. H., Staub, N. C. 1978. Noninvasive measurement of pulmonary transvascular protein flux in sheep. *J. Appl. Physiol.* 45:225–33
61. Gorin, A. B., Kohler, J., DeNardo, G. 1980. Noninvasive measurement of pulmonary transvascular protein flux in normal man. *J. Clin. Invest.* 66:869–77
62. Grimbert, F. A., Parker, J. C., Granger, D. N., Taylor, A. E. 1979. Lung vascular permeability changes following acid aspiration: Effect of altering plasma protein concentration. *Physiologist* 22:4
63. Hackett, P. H., Creagh, C. E., Grover, R. F., Honigman, B. Houston, C. S., Reeves, J. T., Sophocles, A. M., van Hardenbroek, M. 1980. High-altitude pulmonary edema in persons without the right pulmonary artery. *N. Engl. J. Med.* 302:1070–73
64. Hakim, T. S., van der Zee, H., Malik, A. B 1979. Effects of sympathetic nerve stimulation on lung fluid and protein exchange. *J. Appl. Physiol.* 47:1025–30
65. Hales, C. A., Hoop, B., Elmaleh, D., Chesler, D. 1980. Positron tomographic determination of pulmonary vascular permeability. *Am. Rev. Resp. Dis.* 121(Pt. 2):433
66. Hansen, G., Crooks, L. E., Davis, P., DeGroot, J., Herfkens, R., Margulis, A. R., Gooding, C., Kaufman, L., Hoenninger, J., Arakawa, M., McRee, R., Watts, J. 1980. In vivo imaging of rat anatomy with nuclear magnetic resonance. *Radiology* 136:695–700
67. Heflin, A. C., Brigham, K. L. 1979. Granulocyte depletion prevents increased lung vascular permeability after endotoxemia in sheep. *Clin. Res.* 27:399A
68. Henson, P. M., McCarthy, K., Larsen, G. L., Webster, R. O., Gielas, P. C., Dreisin, R. B., King, T. E., Shaw, J. O. 1979. Complement fragments, alveolar macrophages and alveolitis. *Am. J. Pathol.* 97:93–110
69. Hill, J. D., Rattliff, J. L., Pahrott, J. C. W., Lamy, M., Fallat, R. J., Koeniger, E., Yaeger, E. M., Whitmer, G. 1976. Pulmonary pathology in acute respiratory insufficiency: Lung biopsy as a diagnostic tool. *J. Thorac. Cardiovasc. Surg.* 71:64–71
70. Hill, S. L., Elings, V. B., Lewis, F. R. 1980. Changes in lung water and capillary permeability following sepsis and fluid overload. *J. Surg. Res.* 28:140–50
71. Holcroft, J. W., Trunkey, D. D. 1975. Pulmonary extravasation of albumin during and after hemorrhagic shock in baboons. *J. Surg. Res.* 18:91–97
72. Hopewell, P. C., Murray, J. F. 1976. The adult respiratory distress syndrome. *Ann. Rev. Med.* 27:343–56
73. Hultgren, H. N. 1978. High altitude pulmonary edema. In *Lung Water and Solute Exchange*, ed. N. C. Staub, pp. 437–70. New York: Dekker

74. Hurley, J. V. 1978. Current views on mechanisms of pulmonary edema. *J. Pathol.* 125:59–79
75. Ishak, B. A., Seleny, F. L., Noah, Z. L. 1976. Venous air-embolism, a possible cause of acute pulmonary edema. *Anesthesiology* 45:453–55
76. Johnson, A., Malik, A. B. 1980. Effect of granulocytopenia on extravascular lung water content after microembolization. *Am. Rev. Res. Dis.* 122:561–66
77. Jones, J. G., Grossman, R. F., Berry, M., Slavin, G., Hulands, G. H., Minty, B. 1979. Alveolar-capillary membrane permeability. *Am. Rev. Resp. Dis.* 120:399–410
78. Jones, J. G., Lawler, P., Crawley, J. C. W., Minty, B. D., Hulands, G., Veall, N. 1980. Increased alveolar epithelial permeability in cigarette smokers. *Lancet* 1:66–67
79. Kaplan, H. P., Robinson, F. R., Kapanci, Y., Weibel, E. R. 1969. Pathogenesis and reversibility of the pulmonary lesions of oxygen toxicity in monkeys. *Lab. Invest.* 20:94–100
80. Deleted in proof
81. Katz, S., Aberman, A., Frand, U. I., Stein, J. M., Fulop, M. 1972. Heroin pulmonary edema. Evidence for increased pulmonary capillary permeability. *Am. Rev. Resp. Dis.* 106:472–74
82. Kistler, G. S., Caldwell, P. R. B., Weibel, E. R. 1967. Development of fine structural damage to alveolar and capillary lining cells in oxygen-poisoned rat lungs. *Cell Biol.* 32:605–28
83. Kleiner, J. P., Nelson, W. P. 1975. High altitude pulmonary edema—rare disease. *J. Am. Med. Assoc.* 234:491–95
84. Kopaniak, M. M., Issekutz, A. C., Movat, H. Z. 1980. Kinetics of acute inflammation induced by *Escherichia coli* in rabbits—quantitation of blood flow, enhanced vascular permeability, hemorrhage, and leucocyte accumulation. *Am. J. Pathol.* 98:485
85. Landolt, C. C., Matthay, M., Staub, N. C. 1980. Lung resection, increased blood flow and pressure in sheep: failure to detect pore stretching or high linear flow injury. *Microvasc. Res.* 20:117
86. Lewis, F. R., Elings, V. B., Sturm, J. A. 1979. Bedside measurement of lung water. *J. Surg. Res.* 27:250–61
87. Lindquist, P., Saldeen, T., Sandler, H. 1976. Pulmonary damage following pulmonary microembolism in the dog. Effect of various types of treatment. *Acta Clin. Scand.* 142:15–19
88. Lord, P., Olsson, S. E., Audell, L. 1975. Acute pulmonary edema and seizures in hunting dogs. *Nord. Vet. Med.* 27: 112–16
89. Luisada, A. A. 1967. Mechanism of neurogenic pulmonary edema. *Am. J. Cardiol.* 20:66–68
90. Machiedo, G. W., Brown, C. S., Lavigne, J. E., Rush, B. F. 1976. Beneficial effect of prostaglandin-E1 in experimental hemorrhagic shock. *Surg. Gynecol Obstet.* 143:433–36
91. Maldonald, D. 1978. High altitude pulmonary edema. *Radiol. Clin. North Am.* 16:537–49
92. Malik, A. B., Lee, B. C., van der Zee, H., Johnson, A. 1978. Mechanism of neurogenic pulmonary edema. *Am. Rev. Resp. Dis.* 117:367
93. Malik, A. B., Lee, B. C., van der Zee, H., Johnson, A. 1979. Role of fibrin in the genesis of pulmonary edema after embolization in dogs. *Circ. Res.* 45: 120–25
94. Malik, A. B., van der Zee, H. 1978. Mechanism of pulmonary edema induced by microembolization in dogs. *Circ. Res.* 42:72–79
95. Malik, A. B., van der Zee, H., Lee, B. C. 1979. Pulmonary transvascular fluid-dynamics in sheep during hemorrhage. *Lymphology* 12:149–57
96. Matalon, S., McIntyre, B. R., Egan, E. A. 1979. O_2 exposure and permeability of the alveolar epithelium. *Fed. Proc.* 38:1264
97. McIntire, L. V., Dewitz, T. S., Martin, P. R. 1978. Effect of anti-platelet drug dipyridamole (RAF) on leucocyte response to mechancial trauma. *Trans. Am. Soc. Artif. Intern. Organs* 24: 310–14
98. McIntosh, M., Silbergleit, A. 1976. Intravascular platelet and neutrophil aggregation in staging of post-traumatic pulmonary insufficiency. *Surg. Forum* 27:176–77
99. McKeen, C. R., Brigham, K. L., Bowers, R. E., Harris, T. R. 1978. Pulmonary vascular effects of fat emulsion infusion in unanesthetized sheep. *J. Clin. Invest.* 61:1291–97
100. McNamee, J. E., Staub, N. C. 1979. Pore models of sheep lung microvascular barrier using new data on protein tracers. *Microvasc. Res.* 18:229–44
101. Michel, R. P., Hogg, J. C. 1979. The effect of hemorrhagic shock and reinfusion on dog lung. *Fed. Proc.* 38(pt. II):1155
102. Millen, J. E., Glauser, F. L., Smeltzer, D., Egan, P., Propst, K., Fischer, P., Dearden, L., Otis, P. 1978. The role of

leucocytes in ethchlorvynol-induced pulmonary edema. *Chest* 73:75–78

103. Mo Costabella, P., Lindquist, O., Kapanci, Y., Saldeen, T. 1978. Increased vascular permeability in delayed microembolism syndrome—experimental and human findings. *Microvasc. Res.* 15:275–86
104. Modig, J., Hedstrand, U., Fischer, J., Lundstrom, J. 1976. Early recognition and treatment of posttraumatic pulmonary microembolism. *Crit. Care. Med.* 4:180–85
105. Moss, G. S. 1972. Pulmonary involvement in hypovolemic shock. *Ann. Rev. Med.* 23:201–28
106. Myrvold, H. E., Svalande, C. 1977. Pulmonary microembolism in early experimental septic shock—morphological study in dogs. *J. Surg. Res.* 23:65–74
107. Nanjo, S., Bhattacharya, J., Staub, N. C. 1980. Pulmonary edema after HCl aspiration in intact dog lung: effect of albumin therapy. *Physiologist* 23:446
108. Niehaus, G. D., Schumack, P. R., Saba, T. M. 1980. Reticuloendothelial clearance of blood-borne particulates—relevance to experimental lung microembolization and vascular injury. *Ann. Surg.* 191:479–87
109. Northrup,W. F., Humphrey, E. W. 1978. Effect of hemorrhagic shock on pulmonary vascular permeability to plasma proteins. *Surgery* 83:264–73
110. Ogletree, M. L., Bringham, K. L. 1979. Indomethacin augments endotoxin-induced increased lung vascular-permeability in sheep. *Am. Rev. Resp. Dis.* 119:383
111. Ohkuda, K., Nakahara, K., Weidner, W. J., Binder, A., Staub, N. C. 1978. Lung fluid exchange after uneven pulmonary artery obstruction in sheep. *Circ. Res.* 43:152–61
112. Penaloza, D., Sime, F. 1968. Pulmonary circulation during high altitude pulmonary edema. *Bull. Physiopathol. Resp.* 4:17–39
113. Pennock, J. L., Pierce, W. S., Waldhaus, J. A. 1977. Management of lungs during cardiopulmonary bypass. *Surg. Gynecol. Obstet.* 145:917–27
114. Perschau, R. A., Munson, E. S., Chapin, J. C. 1976. Pulmonary interstitial edema after multiple venous air emboli. *Anesthesiology* 45:364–68
115. Peterson, B. T., Ross, J. C., Brigham, K. L. 1980. Naloxone prevention of increased lung vascular permeability with intracranial hypertension in sheep. *Am. Rev. Resp. Dis.* 121(pt. 2):443
116. Petty, T. 1975. The adult respiratory distress syndrome. *Am. Rev. Resp. Dis.* 111:713–15
117. Pietra, G. G., Szidon, J. P., Leventhal, M. M., Fishman, A. P. 1969. Hemoglobin as a tracer in hemodynamic pulmonary edema. *Science* 166:1643–46
118. Pingleton, W. W., Coalson, J. J., Guenter, C. A. 1975. Significance of leucocytes in endotoxic shock. *Exp. Mol. Pathol.* 22:183–94
119. Pistolesi, M., Guintini, C. 1978. Assessment of extravascular lung water. *Radiol. Clin. North Am.* 16:551–74
120. Poddar, P. K., Pierce, E. C. II., Kent, B. 1979. Increased survival in microembolism syndrome (MES) with prostaglandin E_1 (PGE_1) treatment. *Fed. Proc.* 38(pt. II):1266
121. Pratt, P. C. 1965. The reaction of the human lung to enriched oxygen atmosphere. *Ann. NY Acad. Sci.* 121:809–22
122. Prichard, J. S., Lee, G. D. J. 1979. Measurement of water distribution and transcapillary solute flux in dog lung by external radioactivity counting. *Clin. Sci.* 57:145–54
123. Rice, C. L., Doroghaz, P. M., Kohler, J. P., Zarins, C. K., Moss, G. S. 1979. Effect of sepsis and reduced colloid osmotic-pressure on pulmonary-edema. *J. Surg. Res.* 27:347–51
124. Riede, U. N., Joachim, H., Hassensted, J., Costabella, U., Sandritt, W., Augustin, P., Mitterma, C. 1978. Pulmonary air-blood barrier of human shock lungs (clinical, ultrastructural and morphometric study). *Pathol. Res. Pract.* 162: 41–72
125. Rinderknecht, J., Shapiro, L., Krauthammer, M., Taplin, G., Wasserman, K., Uszler, J. M., Effros, R. M. 1980. Accelerated clearance of small solutes from the lungs in interstitial lung disease. *Am. Rev. Resp. Dis.* 121:105–18
126. Robin, E. D., Cross, C. E., Zelis, R. 1973. Pulmonary edema. *N. Engl. J. Med.* 288:239–46, 292–304
127. Sackner, M. A., Greeneltch, D., Heiman, M. S., Epstein, S., Atkins, N. 1975. Diffusing capacity, membrane diffusing capacity, capillary blood volume, pulmonary tissue volume and cardiac output measured by a rebreathing technique. *Am. Rev. Resp. Dis.* 111: 157–65
128. Sacks, T., Moldow, C. F., Craddock, P. R., Bowers, T. K., Jacob, H. S. 1978. Oxygen radicals mediate endothelial cell damage by complement-stimulated granulocytes. *J. Clin. Invest.* 62:1161–67

129. Saldeen, T. 1976. The microembolism syndrome. *Microvasc. Res.* 11:227–59
130. Saldeen, T. 1979. Blood coagulation and shock. *Pathol. Res. Pract.* 165: 221–52
131. Sarkar, T. K., Munshi, A. T. 1977. Postictal pulmonary edema. *Postgrad. Med.* 61:281
132. Sarnoff, S. J., Sarnoff, L. C. 1952. Neurohemodynamics of pulmonary edema. The role of sympathetic pathways in the elevation of pulmonary and systemic vascular pressures following the intracisternal injection of fibrin. *Circulation* 6:51–62
133. Schneeberger, E. E., Karnovsky, M. J. 1971. The influence of intravascular fluid volume on the permeability of newborn and adult mouse lung to ultrastructural protein tracers. *J. Cell Biol.* 49:319–34
134. Scoggin, C. H., Hyers, T. M., Reeves, J. T., Grover, R. F. 1977. High-altitude pulmonary edema in children and young adults of Leadville, Colorado. *N. Eng. J. Med.* 297:1269–72
135. Shanies, H. M. 1977. Noncardiogenic pulmonary edema. *Med. Clin. North Am.* 61:1319–37
136. Shasby, D. M., Fox, R. B., Haradaang, R. N., Repine, J. E. 1980. Mechanisms of pulmonary oxygen toxicity: Neutropenia protects against acute lung injury from hyperoxia. *Am. Rev. Resp. Dis.* 121(pt. 2):258
137. Shellito, J., Murphy, S. 1980. The effect of corticosteroids on rabbit alveolar macrophage function after acid aspiration. *Am. Rev. Resp. Dis.* 121:278
138. Shirley, H. H., Wasserman, K., Wolfram, C. G., Mayerson, H. S. 1957. Capillary permeability to macromolecules: stretched pore phenomenon. *Am. J. Physiol.* 190:189–93
139. Sibbald, W. J., Anderson, R. R., Reid, B., Holliday, R. L., Driedger, A. A. 1980. The effect of high dose corticosteroid therapy on alveoli-capillary permeability in human septic ARDS. *Chest.* In press
140. Simmons, D. H., Nash, G., Brigham, K. L., Tierney, D. F. 1979. Adult respiratory-distress syndrome. *West. J. Med.* 130:218–35
141. Simmons, R. L., Martin, A. M. Jr., Heisterkamp, C. A. III, Ducker, T. B. 1969. Respiratory insufficiency in combat casualties: Pulmonary edema following head injury *Ann. Surg.* 170:39–44
142. Simon, D. S., Murray, J. F., Staub, N. C. 1979. Measurement of pulmonary edema in intact dogs by transthoracic gamma-ray attenuation. *J. App. Physiol.* 47:1228–33
143. Simon, R., Bayne, L. L., Tranbaugh, R. F., Lewis, F. R. 1980. Postictal pulmonary edema: altered capillary permeability outlasts elevated microvascular pressure. *Circulation.* 62:III–58
144. Singh, I., Kapila, C. C., Khanna, P. K., Nanda, R. B., Rao, B. D. P. 1965. High-altitude pulmonary edema. *Lancet* 1:229–34
145. Snashall, P. D. 1980. Pulmonary edema. *Br. J. Dis. Child.* 74:2–22
146. Snashall, P. D., Morgan, B., Keyes, S., McAnulty, R. J., Mitchell-Heggs, P. F. 1978. Radiographic appearances of acute pulmonary edema in dogs related to regional extravascular lung water. *Clin. Sci. Mol. Med.* 54:10P
147. Staub, N. C. 1974. Pulmonary edema. *Physiol. Rev.* 54:678–811
148. Staub, N. C. 1978. Pulmonary edema. Physiologic approaches to management. *Chest* 74:559–64
149. Staub, N. C. 1978. Pulmonary edema due to increased microvascular permeability to fluid and proteins. *Circ. Res.* 43:143–151
150. Staub, N. C. 1980. Pulmonary edema—hypoxia and overperfusion. *N. Eng. J. Med.* 302:1085–86
151. Staub, N. C. 1980. The pathogenesis of pulmonary edema. *Prog. Cardiovasc. Dis.* 23:53–80
152. Staub, N. C. 1980. Mechanism of pulmonary edema following uneven pulmonary artery obstruction and its relationship to high altitude lung injury. In *High Altitude Physiology and Medicine, Vol. II: Distrubances and Therapy,* ed. R. A. Zink, W. Brendel. New York: Springer
153. Staub, N. C., Hogg, J. C. 1980. Conference report: workshop on the measurement of lung water. *Crit. Care. Med.* In press
154. Sturm, J. A., Carpenter, M. A., Lewis, F. R., Graziano, C., Trunkey, D. D. Water and protein movement in the sheep lung water septic shock: effect of colloid versus crystalloid resuscitation. *J. Surg. Res.* 26:233–48
155. Theodore, J., Robin, E. D. 1976. Speculations on neurogenic pulmonary edema. *Am. Rev. Resp. Dis.* 113:405–11
156. Tierney, D. F., Ayers, L., Kasuyama, R. S. 1977. Altered sensitivity to oxygen toxicity. *Am. Rev. Resp. Dis.* 115(Pt. 2): 59–66
157. Todd, T. R. J., Baile, E., Hogg, J. C. 1978. Pulmonary capillary permeability

during hemorrhagic shock. *J. Appl. Physiol.* 45:298–306

158. Toung, T. J. K., Bordos, D., Benson, D. W., Carter, D., Zuidema, G. D., Permutt, S., Cameron, J. L. 1976. Aspiration pneumonia: experimental evaluation of albumin and steroid therapy. *Ann. Surg.* 183:179–84
159. Tranbough, R. F., Lewis, F. R., Christensen, J. M., Elings, V. B. 1980. Lung water changes after thermal injury: the effects of crystalloid resusciation and sepsis. *Ann. Surg.* 192:478–88
160. Deleted in proof
161. Utell, M. J., Wandkte, J. C., Fahey, P. J., Baker, A., Fischer, H. W., Hyde, R. W. 1979. Lung weight in normal and edematous dogs by computerized tomography. *Fed Proc.* 38:1326
162. Vaage, J. 1977. Intravascular platelet aggregation and acute respiratory insufficiency. *Circ. Shock* 4:279–90
163. Vaage, J., Nicolaysen, G., Wasler, B. A. 1976. Aggregation of blood platelets and increased hydraulic conductivity of pulmonary exchange vessels. *Acta Physiol. Scand.* 98:175–84
164. van der Zee, H., Malik, A. B., Lee, B. C., Hakim, T. S. 1980. Lung fluid and protein exchange during intracranical hypertension and role of sympathetic mechanisms. *J. App. Physiol.* 48:273–80
165. Viswanathan, R., Jain, S. K., Subramanian, S. 1969. Pulmonary edema of high altitude. III. Pathogenesis. *Am. Rev. Resp. Dis.* 100:342–49
166. Webb, W. R., Wax, S. D., Kusojima, K. 1974. Microscopic studies of the pulmonary circulation in situ. *Surg. Clin. North Am.* 54:1067–76
167. Wilson, J. W. 1972. Treatment or prevention of pulmonary cellular damage with pharmacologic doses of corticosteroids. *Surg. Gynecol. Obstet.* 134: 675–81
168. Wilson, R. 1973. Acute high-altitude illness in mountaineers. and problems of rescue. *Ann. Intern. Med.* 78:421–28
169. Wolfe, W. G., Devries, W. C. 1975. Oxygen toxicity. *Ann. Rev. Med.* 26:203–17
170. Wray, N. P., Nicotra, M. B. 1978. Pathogenesis of neurogenic pulmonary edema. *Am. Res. Resp. Dis.* 118:783–86

Ann. Rev. Med. 1981. 32:313–26

THE KILLING OF PATHOGENS BY PHAGOCYTES[1]

❖7400

T. G. Gabig, M.D.

Department of Medicine, University of Michigan School of Medicine, and the
Ann Arbor Veterans Administration Hospital, Ann Arbor, Michigan 48105

B. M. Babior, M.D.

Blood Research Laboratory and the Department of Medicine, Tufts–New
England Medical Center, Boston, Massachusetts 02111

Phagocytes include neutrophils, eosinophils, and monocyte-macrophages. They are cells specialized for destruction. Their name reflects the fact that they routinely ingest large particles during the course of their activities (Gr. φαγω, I eat).

The objects that they destroy include invading pathogens, damaged and foreign cells, and connective tissue. Small targets such as bacteria and protozoa are destroyed after phagocytosis, a process by which the target is internalized into a cytoplasmic vesicle (the phagocytic vesicle) lined with what was originally plasma membrane (1). Larger targets can be killed without internalization (2–4). Because many of the destructive agents employed by phagocytes are nonspecifically reactive, normal tissue will be destroyed, especially near a region of inflammation where phagocytes accumulate in large numbers (5–7); this accounts for much of the pathology at sites of inflammation. Pathogens are killed by all three types of phagocytes, though different phagocytes seem to specialize in different types of pathogens. Neutrophils appear to have few other major functions, but eosinophils are thought to limit the extent of an allergic response by the destruction of mediators (8), while the mononuclear phagocytes are the major elements

---

[1] Supported in part by USPHS Grant AI-11827 and a Starter Grant from the Veterans Administration. The US Government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this paper.

involved in the removal of damaged tissue and the elimination of unwanted particulates including foreign bodies, aged red cells, circulating immune complexes, etc. Mononuclear phagocytes (in the form of osteoclasts) are also involved in the remodelling of bone (9).

With few exceptions, the elimination of these objects involves their hydrolytic degradation by enzymes that are present in the lysosomes of the phagocytes. Pathogens, however, are killed before they are degraded. The killing mechanisms on the whole are separate from those involved in degradation, although certain killing mechanisms cause limited degradation of the target as a byproduct of their lethal activity. There is evidence that partial destruction of the outer layers of a pathogen, not fatal per se, may sensitize the organism toward agents that it ordinarily resists. This article is concerned with the mechanisms employed by phagocytes to kill invading pathogens.

Granules and Degranulation

A phagocyte contains lysosome-like granules consisting of proteins and mucopolysaccharides (10, 11) surrounded by a limiting membrane. Many of the proteins in the granules are hydrolytic enzymes (12), serving mainly to degrade ingested targets, while others participate directly in killing by the phagocyte.

Contact between a target and the plasma membrane of the phagocyte leads to a sequence of events known as degranulation. The granules migrate to the region of contact, fuse with the plasma membrane, and release their contents onto the target. The target may be outside the phagocyte but in contact with its membrane, or it may be inside the cell within a phagocytic vesicle.

Recent studies provide tantalizing clues to the mechanism of degranulation, but a complete understanding is some time away. Microtubules facilitate degranulation (13), perhaps by organizing the interior of the cell (14, 15) as they appear to do for chemotaxis (15), but they are not essential for this process (16–18). Ionic movements, including an abrupt efflux of Ca^{2+}, occur in response to agents that also induce degranulation (19); in addition, extracellular ions strongly influence the course of degranulation (18, 20, 21). How these ions exert their effects on granule movement and fusion is not known. Actin-containing microfilaments, the motor of phagocytes, are stripped away from the plasma membrane at sites of degranulation, permitting contact between the granule and the cytoplasmic face of the plasma membrane (22, 23). Are microfilaments also related to granule movement (17, 24)? If so, what is the significance of the finding that the depolymerization of microfilaments by cytochalasin B does not prevent degranulation (17, 21)? Morphological concomitants of granule fusion, including the clearing of membrane-embedded proteins from regions of fusion and the formation of a pentalaminar membrane by the close apposition of the trilaminar granule and plasma membranes, have been reported in several systems (25–27), but a recent study suggests that these may be artifacts of fixation, of uncertain relationship to the fusion process (28). With regard to the biochemical basis of fusion, the recent finding that arachidonic acid is rapidly released from the membrane phospholipids of stimulated neutrophils (29) and mast cells (30) is intriguing in view of the ability of monoacyl phosphatides and free fatty acids to induce cell fusion in other systems by facilitating the reorganization of the phospholipid bilayer (31). These are only two of many types of fusion-promoting agents; others include Ca^{2+}, certain polyols, and the fusogenic glycoproteins of syncytium-forming viruses (32, 33). It remains to be seen whether any of these is involved in the fusion event.

Anaerobic Killing

The occurrence of anaerobic killing mechanisms in phagocytes was demonstrated by Mandell, who showed not only that they exist in these cells but also that against many organisms they are as potent as the oxygen-dependent killing mechanisms (34). Relatively little work has been done, however, on the systems employed by phagocytes to kill in the absence of oxygen. For this reason, anaerobic killing mechanisms are rather less well understood than their aerobic counterparts.

ACIDIFICATION Phagocytic vesicles are rapidly acidified, the pH within the vesicle falling detectably in less than 10 minutes (35). How acidification takes place is not known, though a proton-transporting ATPase would seem possible, since such enzymes occur in other tissues in which pH gradients are generated (36).

The fall in pH potentiates the destruction of the target, since many granule enzymes involved in killing and digestion operate best at pH 4–5 (37). Acidification may therefore be regarded as an auxiliary microbicidal mechanism. By itself, however, acidification has little antimicrobial power. Pneumococci may be killed solely as a result of vesicle acidification, since this species is very susceptible to a drop in the pH of its environment (38). Other organisms are relatively unaffected by a modest fall in the surrounding pH.

HYDROLASES Although most of the hydrolases of phagocytes act only after the target is dead, a few appear to play a direct role in the killing process. These are lysozyme, arginase, and glycosidases.

Lysozyme, found in the granules of all classes of phagocytes and actively secreted by macrophages, is an enzyme that catalyzes the hydrolysis of peptidoglycan, a gigantic macromolecular net which envelops the plasma membrane of a bacterium and provides the organism with its resistance to osmotic lysis (39). Lysozyme acts to destroy this net and strip the organism of its protection against osmotic stress.

Because peptidoglycan is widely distributed among bacteria and fungi, lysozyme should kill a broad spectrum of organisms. Most bacteria, however, are resistant to lysozyme, in some cases because access of the enzyme to the peptidoglycan is prevented by other cell wall constituents (e.g. lipopolysaccharide, in the case of Gram negative organisms), and in other cases because the peptidoglycan itself is altered so as to make it a poor substrate for the enzyme (39). Organisms whose peptidoglycan is susceptible to lysozyme, including *Micrococcus lysodeikticus, Bacillus megatherium* and fungi such as *Cryptococcus neoformans* and *Coccidiodes immitis,* are killed by the enzyme (40–42). Against such organisms, the lysozyme in phagocytes is likely to represent a potent antimicrobial agent. Against most organisms, however, lysozyme is probably of little use, except in certain instances where other systems sensitize intrinsically lysozyme-resistant organisms to the action of the enzyme.

Arginase, an enzyme that catalyzes the hydrolysis of arginine to ornithine and urea, is secreted in large quantities by stimulated macrophages. Certain types of tumor cells are killed when incubated with arginase-secreting macrophages or with medium conditioned by such macrophages, their deaths resulting from arginine starvation due to the destruction of this (for them) essential amino acid by the arginase in the medium (43). Arginase thus represents an oxygen-independent killing agent in this rather artificial system. Whether arginase plays a similar role under more physiological circumstances remains to be established.

Glycosidases are enzymes that cleave polysaccharides. They are important for the degradation of carbohydrate residues in materials ingested by phagocytes (12, 44). Lipopolysaccharides of Gram negative bacteria are largely carbohydrate, and as such may be degraded by these enzymes. Loss of part of the lipopolysaccharide is probably not fatal to the bacteria because mutant organisms lacking portions of the lipopolysaccharide grow well in vitro (45). Such a loss could, however, sensitize an organism to certain microbicidal proteins present in the phagocyte (see below). Thus, glycosidases may serve an important auxiliary role in the lethal action of the phagocyte.

SPECIFIC GRANULE CONTENTS In addition to the azurophil granules (lysosomes) common to all phagocytes, neutrophils and eosinophils each contain a class of specific granules unique to that particular cell. In both cells, the contents of the specific granules are thought to participate in killing.

The specific granules of neutrophils contain apolactoferrin (46), cobalamin-binding protein (47), collagenase (48), and about half the lysozyme in the phagocyte. These substances are liberated into phagocytic vesicles, and in addition are secreted into the extracellular environment in regions of inflammation (49). Two of them—lysozyme and apolactoferrin—participate in killing.

Apolactoferrin is a protein with an exceedingly high affinity for iron (50). It is thought to aid in killing by depriving microorganisms of essential iron. A few organisms are killed outright by iron starvation (51), but most survive, though growth is retarded (52) and virulence is reduced (53). Apolactoferrin thus potentiates microbial killing, but for the most part is not itself microbicidal.

The cobalamin-binding protein is a molecule capable of taking up a wide variety of cobamides and transporting them to the liver for excretion (54). This protein has been postulated to participate in microbial killing, but evidence supporting this role is sparse. An alternative hypothesis is that the protein aids in the disposal of bacterial cobamides, a heterogeneous group of compounds which if not eliminated could act as antimetabolites of cobalamin in mammalian tissues. Though this hypothesis is consistent with what is currently known about the cobalamin-binding protein, it is supported only by indirect evidence. The true function of the protein remains to be established.

The specific granules of eosinophils contain a crystalloid surrounded by an amorphous matrix. The matrix contains a peroxidase different from the myeloperoxidase of neutrophils and mononuclear phagocytes (55, 56) (see above), as well as an acid phosphatase (55). The crystalloid consists largely of major basic protein, a small cationic protein (57).

The major basic protein is involved in host defense against schistosomiasis (57). It acts on the schistosomula, the larval forms that penetrate the skin to initiate an infection. Eosinophils are able to adhere to the schistosomulum, where they discharge the major basic protein onto its surface. This protein, which in pure form can kill the schistosomulum, disrupts the cuticle of the parasite. The eosinophils then burrow into the schistosomulum and the organism is destroyed. Major basic protein also kills trichinella, presumably by a similar mechanism (58).

CATIONIC PROTEINS Cationic proteins are proteins that are positively charged at physiological pH. Many cationic proteins are present in phagocytes, their charge permitting them to interact effectively with the negatively charged surfaces of ingested particles. Phagocytin, a cationic protein extracted from neutrophils by low pH buffers, was the first such to be assigned an antimicrobial function (59). More recently, others have been proposed as antimicrobial agents (60–62). Fractionation reveals that the cationic proteins represent a heterogeneous mixture whose antimicrobial specificity varies from component to component (60–62). Some components of this mixture are proteases, but their proteolytic activity is independent of their antimicrobial function.

These cationic proteins raise many questions. To what extent are different laboratories studying the same proteins? What is the significance of the fact that some of the cationic proteins, at least, are only able to kill microorganisms at unphysiologically low ionic strength (62)? What does the antimicrobial activity of polycations like histone and polyarginine mean with respect to the role of cationic proteins? How do these proteins work?

Answers to several of these questions are suggested by important studies from Elsbach and his collaborators. These workers purified to homogeneity a potent antimicrobial cationic protein from rabbit heterophils (63) and a similar protein from human neutrophils (64). Apart from an antimicrobially irrelevant phospholipase activity present in the rabbit but not the human protein, both of these proteins have similar properties. They are ineffective against Gram positive bacteria, but under physiological conditions of pH and ionic strength they are highly active against Gram negative organisms. *Escherichia coli* mutants whose lipopolysaccharides are incomplete (rough mutants) are much more sensitive than wild-type bacteria to the action of these proteins, but even wild-type organisms are efficiently killed. Killing is a consequence of the binding of the protein to the outer (lipopolysaccharide-containing) membrane of the bacterium (45). The mechanism by which the association between the protein and the outer membrane causes the death of the microorganism is not known.

Oxygen-Dependent Killing

Phagocytes are able to generate powerful oxidizing agents by the partial reduction of oxygen in a metabolic event known as the "respiratory burst." These potent oxidants are employed by the cell to kill its target.

RESPIRATORY BURST This is a change in O_2 metabolism that occurs when phagocytes are exposed to appropriate stimuli. Basal oxygen consumption by unstimulated phagocytes varies with cell type and, in the case of macrophages, with local conditions in the tissue of residence. Unstimulated neutrophils consume little oxygen (65), whereas unstimulated alveolar macrophages consume considerable amounts of oxygen via mitochondrial metabolism (66). All phagocytic cells, however, show large increases over their basal rates of oxygen consumption following exposure to phagocytosable particles or other appropriate stimuli (67–69). Such a stimulated increase in oxygen consumption, which is unaffected by CN^- (67) and therefore has nothing to do with mitochondrial energy production, is accompanied by increases in superoxide (O_2^-) (70–72) and hydrogen peroxide production (73) and in glucose oxidation via the hexosemonophosphate shunt (67).

The biochemical basis for the respiratory burst is thought to be the activation of a membrane-bound oxidase which catalyzes the following reaction:

$$2O_2 + \text{NADPH} \rightarrow 2O_2^- + \text{NADP} \qquad 1.$$

This reaction accounts qualitatively for oxygen consumption and O_2^- production during the burst. The subsequent dismutation of O_2^- to H_2O_2 explains the production of H_2O_2. Dismutation can occur either spontaneously or via

$$2O_2^- + 2H^+ \rightarrow H_2O_2 + O_2 \qquad 2.$$

catalysis by superoxide dismutase, whose concentration in the neutrophil is approximately 10% that of the red cell. Hexosemonophosphate shunt activity is increased by the availability of NADP, which is generated in Reaction 1 and is also produced during detoxification of H_2O_2 by the cytoplasmic glutathione peroxidase-glutathione reductase couple (74, 75):

$$2\ \text{GSH} + H_2O_2 \xrightarrow[\text{glutathione peroxidase}]{} \text{GSSG} + H_2O \qquad 3.$$

$$\text{GSSG} + \text{NADPH} \xrightarrow[\text{glutathione reductase}]{} 2\text{GSH} + \text{NADP} \qquad 4.$$

Activation of the respiratory burst involves a complex series of events that follow the interaction of the stimulus with the plasma membrane of the phagocyte. Activation can be accomplished by many agents, both particulate and soluble (76), and requires neither phagocytosis nor degranulation to take place (77). Energy is necessary for activation (78). Under some circumstances, activation is reversible (77).

Although the molecular basis of activation is poorly understood, the following sequence of events occurs in response to an appropriate stimulus (79):

$$\text{stimulus} \xrightarrow{\text{10 sec}} \begin{array}{c}\text{membrane}\\ \text{depolarization}\end{array} \xrightarrow{\text{30–60 sec}} \left\{\begin{array}{l} O_2^- \text{ production} \\ \Delta\text{cAMP levels} \\ \text{lysosomal enzyme release} \end{array}\right.$$

Membrane depolarization precedes the onset of O_2^- generation, which begins simultaneously with degranulation and changes in intracellular cyclic AMP levels. The nearly simultaneous onset of these three events suggests that neither of the latter two are involved in the activation of the O_2^- generating enzyme.

Recent studies suggest that proteases may be involved in the biochemical pathways leading to burst activation, at least in certain instances. Several synthetic and natural protease inhibitors were found to depress O_2^- generation when the soluble activating agents concanavalin A and cytochalasin E were used as stimuli (80).

Many workers believe that the enzyme which catalyzes Reaction 1 is an extensively studied membrane-associated (81) flavoprotein (82) which can use either NADPH or NADH as electron donor, but prefers the former (83, 84). This enzyme can be detected in preparations from normal phagocytes that have been exposed to a specific activating stimulus prior to disruption. But it is absent in preparations obtained from resting neutrophils or from stimulated neutrophils isolated from patients with chronic granulomatous disease (CGD), an inherited condition in which phagocytes are unable to express a respiratory burst (see below). Some experts, however, believe that other enzymes, particularly an NADH-specific oxidase, may be responsible for O_2^- formation during the respiratory burst (reviewed in 85 and 86). A set of useful criteria for identifying the O_2^--forming enzyme has been proposed to help settle the issue (87).

The possibility that a b-type cytochrome participates in the respiratory burst has recently attracted interest. A heme protein with a cytochrome b-like spectrum was discovered in rabbit neutrophils in 1966 (88), and in human cells in 1977 (89). The cytochrome is absent in some patients with CGD, which suggests a connection with the respiratory burst, but other CGD patients have normal amounts in their neutrophils (90). Questions have been raised as to whether this protein might represent myeloperoxidase, eosinophil peroxidase, or contaminating microsomal cytochrome b_5 or P_{450}. The cytochrome can be separated from myeloperoxidase in subcellular fractionation studies of neutrophil homogenates (91), which rules out identity with myeloperoxidase, but convincing data that this cytochrome is not a microsomal cytochrome b_5 or altered P_{450} (i.e. P_{420}) is lacking. More importantly, there is no clear evidence that this cytochrome is involved in the oxidase activity of the respiratory burst (i.e. Reaction 1 above). At present the function of this b-type cytochrome of human neutrophils is unclear, as is its relation to the respiratory burst oxidase.

KILLING Several O_2-dependent killing mechanisms are postulated to occur in phagocytes. Two are on reasonably solid experimental footing: the peroxide-halide-myeloperoxidase mechanism (92), and a mechanism employing highly reactive oxidizing radicals. Other proposed mechanisms have little experimental support as of now.

Myeloperoxidase is a heme enzyme present in high concentration in the azurophil granules (93). It is a constituent of a microbicidal system whose other components are H_2O_2 and Cl^-. These serve as substrates for the peroxidase, which catalyzes their conversion to HOCl and water (94). The HOCl is thought to be the proximate microbicidal agent in this system. How it kills the microorganism is not known, though the findings that the system can chlorinate intact bacteria (95) and can mediate the oxidation of amino and sulfhydryl groups of proteins (96) suggest indirectly that killing may result through the alteration of essential macromolecules.

Phagocytes have been shown to produce highly reactive oxidizing radicals, including hydroxyl radical ($OH\cdot$) and possibly others (97, 98, 98a,b). These radicals arise by the reaction of O_2^- and H_2O_2 (99) and perhaps other peroxy compounds (100). The reaction between O_2^- and H_2O_2 is known as the Haber-Weiss reaction; it requires iron as a catalyst.

Reactive oxidizing radicals can kill bacteria both in artificial systems (101) and in neutrophils themselves (102). Oxidizing radicals probably represent the proximate effectors in the oxygen-requiring myeloperoxidase-independent microbicidal systems of neutrophils. Nothing is known about how they kill bacteria, though their ability to damage DNA suggests that they may act by destroying the genome.

Other agents that have been postulated as effectors in oxygen-dependent killing by neutrophils are H_2O_2, O_2^- itself and singlet oxygen. H_2O_2 alone is weakly microbicidal and may act in this capacity in neutrophils, particularly in combination with ascorbic acid (103). O_2^-, on the other hand, is surprisingly innocuous and appears to play no role in bacterial killing by neutrophils except as a precursor of more potent oxidants. Singlet oxygen, a reactive electronically excited form of O_2, has been postulated as a microbicidal agent primarily on the basis of experiments with singlet oxygen trapping reagents (104) and studies of neutrophil chemiluminescence (105). The evidence is ambiguous, however, because neither light emission nor the trapping reagents are specific for singlet oxygen (106, 107). Furthermore, theoretical considerations suggest that singlet oxygen, if formed, is unlikely to survive long enough to do much damage (108). We believe that singlet oxygen has little if anything to do with microbial killing by neutrophils.

Inherited Abnormalities Affecting Microbicidal Systems

Defects in microbial killing that are well characterized from a biochemical point of view have not yet been reported among the anaerobic killing systems, but several involving oxygen-dependent killing have been described.

CHRONIC GRANULOMATOUS DISEASE CGD, discovered by Berendes et al (109), was the first inherited abnormality of bacterial killing to be described. The disease is generally characterized by recurrent and severe bacterial infections from early childhood, though some affected individuals have little trouble. The functional defect is an inability of the cells to express a respiratory burst (110, 111). CGD neutrophils thus cannot generate microbicidal oxidants and have to rely exclusively on anerobic mechanisms for the destruction of most invading pathogens. The bacteria most resistant to these mechanisms are *S. aureus* and the enterobacteriaceae. These bacteria are killed poorly by CGD neutrophils, and it is these organisms that are responsible for most of the infections in patients with CGD (112).

Biochemically, the disease is characterized by an inability to detect the activity of the O_2^--forming oxidase in cell-free preparations obtained from stimulated CGD neutrophils (113). In cases studied to date, the lack of oxidase activity appears to be due to a defect in the apparatus responsible for activating the dormant oxidase (114, 114a); whether or not the oxidase itself is present in CGD neutrophils is not known. Indeed, two cases have been reported in which O_2^--forming activity is elicited with some normally competent stimuli but not with others (115, 116), which suggests that the activation apparatus itself consists of multiple elements, malfunction of any of which will give rise to CGD.

CGD neutrophils have been used as an experimental tool to aid in the assignment of the respiratory burst oxidase. As of now there are three candidates vying for this honor: the membrane-associated NADPH oxidase; an NADH oxidase whose location in the neutrophil may vary with the cell's metabolic state; and a particulate b-type cytochrome. These were discussed above. In our view, the present evidence strongly supports the membrane-associated NADPH oxidase as the enzyme responsible for the respiratory burst.

CGD is easy to diagnose. The major symptom complex suggesting its presence is a history of recurrent deep bacterial infections in a child or an adult. But it is difficult to manage (116a). Prophylactic antibiotics have partially ameliorated the problems posed by acute bacterial infections (117), but fungal infections are now a serious problem. In addition, there is a rising incidence of fibrosis with secondary organ dysfunction, presumably a result of imperfect suppression of localized infections with persistent chronic inflammation. In this category of complications, pyloric stricture, bowel obstruction, bladder fibrosis, and pulmonary fibrosis have all been described. Approaches to the prevention of these complications might include yet-to-be described improvements in antibiotic prophylaxis, bone marrow transplantation (118), and genetic counselling. This last approach has recently been strengthened by the development of a test to diagnose CGD in the fetus (119).

GLUCOSE-6-PHOSPHATE DEHYDROGENASE (G-6-PD) DEFICIENCY Though customarily regarded as a disease of red cells, G-6-PD deficiency, if severe enough, leads to an impairment in the respiratory burst of neutrophils, with a clinical picture very similar to CGD except for the hemolytic anemia (120, 121). The failure of neutrophils to express a respiratory burst in this condition is presumably due to inefficient generation of NADPH, the reducing agent used by the respiratory burst oxidase. The impairment of the respiratory burst in severe G-6-PD deficiency is an additional piece of evidence supporting NADPH oxidase, as opposed to NADH oxidase, as the enzyme responsible for the respiratory burst.

MYELOPEROXIDASE (MPO) DEFICIENCY MPO deficiency, originally thought to be an unusual condition, is apparently rather common, though rarely suspected because it produces few symptoms (122). This is because the destruction of bacteria by MPO-deficient cells, while greatly delayed, is eventually complete, a situation contrasting with CGD, in which bacteria survive and multiply with affected neutrophils (123). Oxidizing radicals apparently fulfill the antibacterial role of the peroxide-halide-MPO system in MPO-deficient cells. On the other hand, *C. albicans* survive within these cells (124). It is of interest that while patients with MPO deficiency are normally resistant to bacterial infections, they seem to be unusually susceptible to candidiasis, particularly if they are diabetic.

DEFICIENCY OF H_2O_2-DESTROYING ENZYMES Neutrophils defend themselves against destruction by their own oxidants by means of four enzymes comprising three antioxidant systems: (*a*) superoxide dismutase, which converts O_2^- to oxygen and H_2O_2; (*b*) glutathione reductase plus glutathione peroxidase, which together reduce H_2O_2 to water at the expense of NADPH (see above); and (*c*) catalase, which dismutes H_2O_2 to oxygen and water. The glutathione-dependent system is equipped to handle low concentrations of H_2O_2, while catalase is most effective at higher concentrations (125).

Familial glutathione reductase deficiency was described by Roos and co-workers (126). Inherited as an autosomal recessive condition, the disease presents as a hemolytic anemia brought on by oxidant stress. The respiratory burst of reductase-deficient neutrophils is much briefer than normal, lasting for only 3–5 minutes instead of the usual 20–30. The biochemical basis for this abbreviated burst is not yet worked out, but it presumably has to do with damage to the neutrophil brought about by either depletion of reduced gluthathione or the unopposed action of the H_2O_2 that diffuses from its site of origin into the cytoplasm of the cell. In any case, the truncated burst is sufficient, since patients with glutathione reductase deficiency resist infection normally, and *S. aureus* are killed normally by glutathione reductase-deficient neutrophils.

The same laboratory studied neutrophils from patients with acatalasia (127). Left to themselves, cells from these patients behave normally. When exposed to external H_2O_2, however, they acquire defects in degranulation and the respiratory burst that are not seen in similarly treated normal neutrophils. The clinical benignity of acatalasia indicates that the defect in neutrophil function induced by exposure to H_2O_2 is of no functional significance.

Literature Cited

1. Hirsch, J. G. 1962. *J. Exp. Med.* 116:827–34
2. Anwar, A. R. E., Smithers, S. R., Kay, A. B. 1979. *J. Immunol.* 122:628–37
3. Bass, D. A., Szejda, P. 1979. *J. Clin. Invest.* 64:1415–22
4. Nathan, C. F., Brukner, L. H., Silverstein, S. C., Cohn, Z. A. 1979. *J. Exp. Med.* 149:84–99
5. Janoff, A., Feinstein, G., Malemud, C. L., Elias, J. M. 1976. *J. Clin. Invest.* 57:615–24
6. Lieberman, J. 1976. *Chest* 70:62–67
7. Sacks, T., Moldow, C. F., Craddock, P. R., Bowers, T. K., Jacobs, H. S. 1978. *J. Clin. Invest.* 61:1161–67
8. Kay, A. B. 1976. *Br. J. Haematol.* 33:313–18
9. Ash, P., Loutit, J. F., Townsend, K. M. S. 1980. *Nature* 283:669–70
10. Olsson, I. 1969. *Biochim. Biophys. Acta* 177:241–49
11. Avila, J. L., Convit, J. 1976. *Biochem. J.* 160:129–36
12. Bretz, U., Baggiolini, M. 1974. *J. Cell Biol.* 63:251–69
13. Goldstein, I., Hoffstein, S., Gallin, J., Weissmann, G. 1973. *Proc. Natl. Acad. Sci. USA* 70:2916–20
14. Schliwa, M. 1978. *J. Cell Biol.* 76:605–14
15. Malech, H. L., Root, R. K., Gallin, J. I. 1977. *J. Cell Biol.* 75:666–93
16. Hoffstein, S., Goldstein, I. M., Weissmann, G. 1977. *J. Cell Biol.* 73:242–56
17. Hoffstein, S., Weissmann, G. 1978. *J. Cell Biol.* 78:769–81
18. Becker, E. L., Showell, H. J. 1974. *J. Immunol.* 112:2055–62
19. Naccache, P. H., Showell, H. J., Becker, E. L., Sha'afi, R. I. 1977. *J. Cell Biol.* 75:635–49
20. Goldstein, I. M., Hoffstein, S. T., Weissmann, G. 1975. *J. Immunol.* 115:665–70
21. Showell, H. J., Naccache, P. H., Sha'afi, R. I., Becker, E. L. 1977. *J. Immunol.* 119:804–11
22. Stendahl, O. I., Hartwig, J. H., Brotschi, E. A., Stossel, T. P. 1980. *J. Cell Biol.* 84:215–24
23. Berlin, R. D., Oliver, J. M. 1978. *J. Cell Biol.* 77:789–804
24. Moore, P. L., Bank, H. L., Spicer, S. S. 1976. *J. Cell Biol.* 71:659–66
25. Chi, E. Y., Lagunoff, D., Koehler, J. K. 1976. *Proc. Natl. Acad. Sci. USA* 73:2823–27
26. Moore, P. L., Bank, H. L., Brissie, N. T., Spicer, S. S. 1978. *J. Cell Biol.* 76:158–74
27. Tanaka, Y., de Camilli, P., Meldolesi, J. 1980. *J. Cell Biol.* 84:438–53
28. Chandler, D. E., Heuser, J. 1979. *J. Cell Biol.* 83:91–108
29. Waits, M., DeChatelet, L. R., King, L., Shirley, P. S. 1979. *Biochem. Biophys. Res. Commun.* 90:984–92
30. Kennerly, D. A., Sullivan, T. J., Sylwester, P., Parker, C. W. 1979. *J. Exp. Med.* 150:1039–44
31. Kantor, H. L., Prestegard, J. H. 1975. *Biochemistry* 14:1790–95
32. Ahkong, Q. F., Fisher, D., Tampion, W., Lucy, J. A. 1973. *Biochem. J.* 136:147–55
33. Scheid, A., Choppin, P. W. 1974. *Virology* 57:475–90
34. Mandell, G. L. 1974. *Infect. Immun.* 9:337–41
35. Jacques, Y. V., Bainton, D. F. 1978. *Lab. Invest.* 39:179–85
36. Berglindh, T., Dibona, D. R., Pace, C. S., Sachs, G. 1980. *J. Cell Biol.* 85:392–401
37. Avila, J. L., Convit, J. 1973. *Biochim. Biophys. Acta* 293:397–408
38. Lord, F. T., Nye, R. N. 1922. *J. Exp. Med.* 35:685–87
39. Strominger, J. L., Ghuysen, J. M. 1967. *Science* 158:213–21
40. Brumfitt, W., Glynn, A. A. 1961. *Br. J. Exp. Pathol.* 42:408–23
41. Gadebusch, H. H., Hohnson, A. G. 1966. *J. Infect. Dis.* 116:551–65
42. Collins, M. S., Pappagianis, D. 1974. *Infect. Immun.* 10:616–23

43. Currie, G. A. 1978. *Nature* 273:758–59
44. Rest, R. F., Cooney, M. H., Spitznagel, J. K. 1978. *Biochem. J.* 174:53–59
45. Weiss, J., Beckerdite-Quagliata, S., Elsbach, P. 1980. *J. Clin. Invest.* 65:619–28
46. Leffell, M. S., Spitznagel, J. K. 1972. *Infect. Immun.* 6:761–65
47. Kane, S. P., Peters, T. J. 1975. *Clin. Sci. Mol. Med.* 49:171–82
48. Murphy, G., Reynolds, J. J., Bretz, U., Baggiolini, M. 1977. *Biochem. J.* 162:195–97
49. Wright, D. G., Gallin, J. I. 1979. *J. Immunol.* 123:285–94
50. Masson, P. L., Heremans, J. F. 1968. *Eur. J. Biochem.* 6:579–84
51. Arnold, R. R., Cole, M. F., McGhee, J. R. 1977. *Science* 197:263–65
52. Oram, J. D., Reiter, B. 1968. *Biochim. Biophys. Acta* 170:351–65
53. Payne, S. M., Finkelstein, R. A. 1978. *J. Clin. Invest.* 61:1428–40
54. Kolhouse, J. F., Allen, R. H. 1977. *J. Clin. Invest.* 60:1381–92
55. Heyneman, R. A., Monbaliu-Bauwens, D., Vercauteren, R. E. 1975. *Comp. Biochem. Physiol.* 50B:463–69
56. Salmon, S. E., Cline, M. J., Schultz, J., Lehrer, R. I. 1970. *N. Engl. J. Med.* 282:250–53
57. Butterworth, A. E., Wassom, D. L., Gleich, G. J., Loegering, D. A., David, J. R. 1979. *J. Immunol.* 122:221–29
58. Kazura, J. W., Aikawa, M. 1980. *J. Immunol.* 124:355–61
59. Hirsch, J. G. 1956. *J. Exp. Med.* 103:589–611
60. Zeya, H. I., Spitznagel, J. K. 1966. *J. Bacteriol.* 91:755–62
61. Lehrer, R. I. 1975. *J. Clin. Invest.* 55:338–46
62. Odeberg, H., Olsson, I. 1975. *J. Clin. Invest.* 56:1118–24
63. Weiss, J., Franson, R. C., Beckerdite, S., Schmiedler, K., Elsbach, P. 1975. *J. Clin. Invest.* 55:33–42
64. Weiss, J., Elsbach, P., Olsson, I., Odeberg, H. 1978. *J. Biol. Chem.* 253:2664–72
65. Karnovsky, M. L. 1968. *Semin. Hematol.* 5:156–65
66. Simon, L. M., Robin, E. D., Phillips, J. R. 1977. *J. Clin. Invest.* 59:443–48
67. Sbarra, A. J., Karnovsky, M. L. 1959. *J. Biol. Chem.* 234:1355–62
68. Oren, R., Farnham, A. E., Saito, K. 1963. *J. Cell Biol.* 17:487–501
69. Baehner, R. L., Johnston, R. B. Jr. 1971. *Br. J. Haematol.* 20:277–85
70. Babior, B. M., Kipnes, R. S., Curnutte, J. T. 1973. *J. Clin. Invest.* 52:741–44
71. Sagone, A. L. Jr., King, G. W., Metz, E. N. 1976. *J. Clin. Invest.* 57:1352–58
72. Klebanoff, S. J., Durack, D. T., Rosen, H. 1977. *Infect. Immun.* 17:167–73
73. Root, R. K., Metcalf, J. A. 1977. *J. Clin. Invest.* 60:1266–79
74. Reed, P. W. 1969. *J. Biol. Chem.* 244:2459–64
75. Voetman, A. A., Loos, J. A., Roos, D. 1980. *Blood* 55:741–47
76. Cheson, B. D., Curnutte, J. T., Babior, B. M. 1977. *Prog. Clin. Immunol.* 3:1–65
77. Curnutte, J. T., Babior, B. M., Karnovsky, M. L. 1979. *J. Clin. Invest.* 63:637–47
78. Cohen, H. J., Chovaniec, M. E. 1978. *J. Clin. Invest.* 61:1088–96
79. Korchak, H. M., Weissmann, G. 1978. *Proc. Natl. Acad. Sci. USA* 75:3818–22
80. Kitagawa, S., Takaku, F., Sakamoto, S. 1980. *J. Clin. Invest.* 65:74–81
81. Dewald, B., Baggiolini, M., Curnutte, J. T., Babior, B. M. 1979. *J. Clin. Invest.* 63:21–29
82. Babior, B. M., Kipnes, R. S. 1977. *Blood* 50:517–24
83. Babior, B. M., Curnutte, J. T., McMurrich, B. J. 1976. *J. Clin. Invest.* 58:989–96
84. Gabig, T. G., Babior, B. M. 1979. *J. Biol. Chem.* 254:9070–74
85. Babior, B. M. 1978. *N. Engl. J. Med.* 298:659–68, 721–25
86. Badwey, J. A., Karnovsky, M. L. 1980. *Ann. Rev. Biochem.* 49:695–726
87. Badwey, J. A., Curnutte, J. T., Karnovsky, M. L. 1979. *N. Engl. J. Med.* 300:1157–60
88. Shinagawa, Y., Tanaka, C., Teraoka, A., Shinagawa, Y. 1966. *J. Biochem.* 59:622–24
89. Segal, A. W., Webster, D., Jones, O. T. G., Allison, A. C. 1978. *Lancet* 2:446–49
90. Borregaard, N., Johansen, K., Tandorff, E., Wandall, J. H. 1979. *Lancet* 1:949–51
91. Segal, A. W., Jones, O. T. G. 1979. *Biochem. J.* 182:181–88
92. Klebanoff, S. J. 1968. *J. Bacteriol.* 95:2131–38
93. Bainton, D. F., Farquhar, M. G. 1968. *J. Cell Biol.* 39:299–317
94. Harrison, J. E., Schultz, J. 1976. *J. Biol. Chem.* 251:1371–74
95. Zglicyznski, J. M., Stelmaszynska, T. 1975. *Eur. J. Biochem.* 56:157–62
96. Thomas, E. L. 1979. *Infect. Immun.* 23:522–31
97. Tauber, A. I., Babior, B. M. 1977. *J. Clin. Invest.* 60:374–79

98. Weiss, S. J., King, G. W., LoBuglio, A. F. 1978. *J. Clin. Invest.* 60:370–73
98a. Green, M. R., Hill, H. A. O., Okolow-Zubkowska, M. J., Segal, A. W. 1979. *FEBS Lett.* 100:23–26
98b. Rosen, H., Klebanoff, S. J. 1979. *J. Clin. Invest.* 64:1725–29
99. McCord, J. M., Day, E. D. 1978. *FEBS Lett.* 92:321–26
100. Peters, J. W., Foote, C. S. 1976. *J. Am. Chem. Soc.* 98:873–75
101. Rosen, H., Klebanoff, S. J. 1979. *J. Exp. Med.* 149:27–39
102. Johnston, R. B., Keele, B. B., Misra, H. P., Lehmeyer, J. E., Webb, L. S. Baehner, R. L., Rajagopalan, K. V. 1975. *J. Clin. Invest.* 55:1357–73
103. Drath, D. B., Karnovsky, M. L. 1974. *Infect. Immun.* 10:1077–83
104. Rosen, H., Klebanoff, S. J. 1977. *J. Biol. Chem.* 252:4803–10
105. Allen, R. C., Stjernholm, R. L., Steele, R. H. 1972. *Biochem, Biophys. Res. Commun.* 47:679–84
106. Held, A. M., Hurst, J. K. 1978. *Biochem. Biophys. Res. Commun.* 81:878–85
107. Cheson, B. D., Christensen, R. L., Sperling, R., Kohler, B. E., Babior, B. M. 1976. *J. Clin. Invest.* 58:789–96
108. Merkel, P. B., Kearns, D. R. 1972. *J. Am. Chem. Soc.* 94:1029–30
109. Berendes, H. Bridges, R. A., Good, R. A. 1957. *Minn. Med.* 40:309–12
110. Holmes, B., Page, A. R., Good, R. A. 1967. *J. Clin. Invest.* 46:1422–32
111. Baehner, R. L., Nathan, D. G. 1967. *Science* 155:835–36
112. Johnston, R. B. Jr., Baehner, R. L. 1971. *Pediatrics* 48:730–39
113. Curnutte, J. T., Kipnes, R. S., Babior, B. M. 1975. *N. Engl. J. Med.* 293:628–32
114. Whitin, J. C., Chapman, C. E., Simons, E. R., Chovaniec, M. E., Cohen, H. J. 1980. *J. Biol. Chem.* 255:1874–78
114a. Seligmann, B. E., Gallin, J. I. 1980. *J. Clin. Invest.* 66:493–503
115. Weening, R. S., Roos, D., Weemaes, C. M. R., Homan-Muller, J. W. T., van Schaik, M. L. J. 1976. *J. Lab. Clin. Med.* 88:757–68
116. Harvath, L., Anderson, B. R. 1979. *N. Engl. J. Med.* 300:1130–35
116a. Johnston, R. B. Jr., Newman, S. L. 1977. *Pediatr. Clin. North Am.* 24: 365–76
117. Philippart, A. I., Colodny, A. H., Baehner, R. L. 1972. *Pediatrics* 50:923–25
118. Westminister Hospitals Bone-Marrow Transplant Team. 1977. *Lancet* 1: 210–13
119. Newburger, P. E., Cohen, H. J., Rothchild, S. B., Hobbins, J. C., Malawista, S. E., Mahoney, M. J. 1979. *N. Engl. J. Med.* 300:178–81
120. Baehner, R. L., Johnston, R. B. Jr., Nathan, D. G. 1972. *J. Reticuloendothel. Soc.* 12:150–69
121. Gray, G. R., Klebanoff, S. J., Stamatoyannopoulos, G. 1973. *Lancet* 2: 530–34
122. Kitahara, M., Simonian, Y., Eyre, H. J. 1979. *J. Lab. Clin. Med.* 93:232–37
123. Klebanoff, S. J. 1970. *Science* 169: 1095–97
124. Lehrer, R. I., Cline, M. J. 1969. *J. Clin. Invest.* 48:1478–88
125. Nicholls, P., Schonbaum, G. R. 1963. *The enzymes,* ed. P. D. Boyer, H. Lardy, K. Myrback, pp. 198–200. New York: Academic. 2nd ed.
126. Roos, D., Weening, R. S., Voetman, A. A., van Schaik, M. L. J., Bot, A. A. M., Meerhof, L. J., Loos, J. A. 1979. *Blood* 53:851–66
127. Roos, D., Weening, R. S., Wyss, S. R., Aebi, H. E. 1980. *J. Clin. Invest.* 65:1515–22

Ann. Rev. Med. 1981. 32:327–33
Copyright © 1981 by Annual Reviews Inc. All rights reserved

CLOSTRIDIA AND ANTIMICROBIAL ENTEROCOLITIS

❖7401

Joseph Silva, Jr., M.D. and Robert Fekety, M.D.

Department of Internal Medicine and Division of Infectious Diseases,
University of Michigan, Ann Arbor, Michigan 48109

Introduction

Pseudomembranous enterocolitis has been associated with gastrointestinal surgery, low birth weight, intestinal obstruction, uremia, hypotensive cardiovascular diseases, and drugs (1). Microscopically, inflammation is seen in the superficial colonic mucosa along with mucosal ulcerations and the formation of a pseudomembrane composed of fibrinogen, degenerating mucosal cells, and leukocytes. During the past decade, antibiotic therapy has been the major cause of pseudomembranous colitis in man, leading to use of the term antibiotic-associated colitis (AAC). In the mid 1970s, AAC was frequently attributed to clindamycin and lincomycin therapy, but since then most antibiotics have been implicated in its causation. The direct cause of AAC is not an antibiotic but toxigenic *Clostridium difficile*, a large gram-positive anaerobic bacillus originally isolated in 1935 from the feces of 4 of 10 healthy infants (2). Since then, *C. difficile* has occasionally been recovered from patients with septicemia or wound infections and is now recognized as playing an important role in enterocolitic inflammatory diseases.

Role of Clostridia in Diarrhea and Colitis

Clostridial toxins have diverse biological characteristics such as dermonecrotizing, hemolytic, edema-producing, neurotoxic, cardiotoxic, enterotoxic, and other proteolytic activities. Various domesticated animals have enzootic and epizootic syndromes of diarrhea that are frequently caused by *C. perfringens* and other clostridia. In humans, *C. perfringens* (type A) only causes a watery diarrhea while toxins of *C. perfringens* (type C) in some countries causes "pig-bel," a life-threatening disease of abdomi-

0066-4219/81/0401-0327$01.00

nal bloating and hemorrhagic diarrhea. Pig-bel usually follows pork consumption by individuals who usually live on a protein-deficient diet. Exciting work from Papua New Guinea has shown that a toxoid of this clostridial toxin can prevent "pig-bel" in humans (3), a finding that may be applicable to other clostridial diarrheas.

The etiological role of *C. difficile* in AAC was discovered by several groups between 1976 and 1978 employing the hamster model. This animal was used by Small in 1968 to study the gastrointestinal effects of lincomycin; however, his data gained little attention because few reports of serious diarrhea related to this drug existed then (4). Subsequently it was shown in hamsters that a single, low dose of clindamycin caused a hemorrhagic cecitis that resembled AAC in humans (5, 6)

After hamsters were given an antibiotic challenge, some clostridial genera (particularly *C. difficile*) increased quantitatively and a fecal cytotoxin appeared. Most antibiotics tested in the hamster cause similar lesions (7). Colitis was also produced in hamsters by giving intracecally, or orgastrically, large numbers of *C. difficile* grown in broth cultures (5). Bacteria-free fecal filtrates from antibiotic-treated animals with colitis contained a cytotoxin that caused AAC after being given intracecally or orgastrically to normal animals. Similar biological effects were seen when broth filtrates of *C. difficile,* or fecal toxins from diseased animals with colitis were given intradermally to rabbits, or intracecally or intraperitoneally to hamsters. The toxin(s) of *C. difficile* are neutralized in vitro by incubation with an antitoxin raised against *C. sordellii* toxin (8). Current thoughts are that immunological similarities probably exist between toxins of *C. difficile* and *C. sordellii.* Hamsters immunized passively with *C. sordellii* antitoxin by several routes survived lethal challenges of intracecally administered *C. difficile* toxin or a lethal challenge of antibiotic (9). Koch's postulates were therefore fulfilled in the hamster and data generated in humans closely paralleled these studies (10, 11).

C. difficile may produce several types of toxins and most isolates are toxigenic. To date, the best-characterized toxin causes cytopathic alterations in tissue culture monolayers of fibroblastic lines. Small amounts (ng/ml) of this cytopathic toxin lead to a retraction of cultured cells, which then "lift-off" from the container surface. *C. difficile* cytotoxin is heat-labile and susceptible to acid, alkaline, and trypsin treatment. It is not ether-extractable and has a molecular weight $> 1 \times 10^5$. Following ammonium sulfate precipitation, the toxin appears to be protein with an affinity for anionic-exchange resins (12). This assay has been proven by many groups to be positive in confirmed cases of AAC. For instance, Bartlett et al (13) found that stools of 75 of 76 patients with confirmed enterocolitis contained the toxin. *C. difficile* toxin was not found in feces of patients with other forms of gastrointestinal disease. Similarly this assay has usually been negative in patients who are taking antibiotics and have watery diarrhea (without biopsy-proven colitis.)

Incidence and Epidemiology of C. difficile *Colitis*

The incidence of antimicrobial and clostridial enterocolitis throughout the world is not well delineated. Many recent reports have added to the list of antibiotics that cause AAC, and only a few antibiotics are free from this indictment. AAC is caused by antibiotics to which *C. difficile* may be either sensitive or resistant by in vitro susceptibility testing. The best incidence data were accrued when much attention was focused on clindamycin in 1977–1978. A letter to physicians from the manufacturer indicated an overall incidence in the US of 1 per 100,000 treated patients. In Dallas, the incidence of AAC for clindamycin was 1 per 5000 whereas in St. Louis, diarrhea was seen in 21% of patients who received clindamycin, and colitis in 10% (1). Other studies from Louisiana and Ohio revealed no significant diarrhea or enterocolitis with clindamycin. These incidences may represent regional differences in occurrence of *C. difficile* colonization within the population or differences in diagnostic criteria used by various physicians.

The human and/or environmental reservoirs of *C. difficile* have not been well studied. Thus far, *C. difficile* has been isolated from the urogenital cultures of 72% of 108 women in a venereal disease clinic and all 42 men cultured with nonspecific urethritis (17). Other studies are underway now that an excellent selective media is available.

Newer roles for *C. difficile* in causation of colonic disease have recently been described. The organism and its toxin were found in feces of patients having exacerbations of chronic ulcerative colitis (14) and Crohn's disease (15), and in the diarrheal stools of a few patients with pseudomembranous enterocolitis following the use of antineoplastic drugs but not antibiotics; the latter patients responded satisfactorily to treatment with oral vancomycin (16).

Clinical Manifestations of Antibiotic-Associated Colitis

Clostridial enterocolitis is seen in all ages but is more common in elderly women. Symptoms generally begin within a week of starting antibiotic therapy but may occur up to four weeks after stopping antibiotic(s). AAC is seen more commonly with oral administration of antibiotics but does occur with parenteral therapy. While the degree of diarrhea is variable, most patients with biopsy-proven enterocolitis have a watery diarrhea numbering 10–20 watery stools per day. Over half of our patients with AAC have blood and/or mucous in their stools. Abdominal tenderness is usually minimal but occasional patients will have a "surgical" abdomen with abdominal rigidity and toxic megacolon. Fever and a hypoalbuminemia are also seen in the majority of patients with AAC, and a peripheral blood leukocytosis is common (18).

Diagnosis

AAC is diagnosed by only a few tests. Examination for fecal leukocytes has not been helpful nor have gram stains to identify large numbers of gram-positive rods. Stools should be examined for ova and parasites and cultured selectively for other enteric pathogens (such as salmonella or shigella) because other pathogens can masquerade as clostridial enterocolitis. A selective agar media incorporating antibiotics (cycloserine and cefoxitin) is now available and can detect as few as 100 organisms of *C. difficile* per gram (wet weight) of feces (19). Unfortunately cultural identification requires 2–3 days of incubation under fastidious anaerobic conditions, and some patients may carry *C. difficile* without toxin being present in their stools.

One useful assay is to identify *C. difficile* toxin by incubation of bacteria-free filtrates of patients' stools with fibroblastic tissue cultures. Toxic stool filtrates cause characteristic changes within 6–24 hours of incubation in these tissue cultures. However, some bacterial toxins in fecal filtrates may be missed by these culture lines.

Batts et al (20) found stool filtrates from three patients with AAC that caused cytotoxicity only in a rat mammary carcinoma cell culture (Walker line). AAC in these patients and the cytopathogenicity of their feces was attributed to toxigenic *Staphylococcus aureus.* In the early 1960s, enterotoxic *S. aureus* were felt to be the major cause of AAC but this idea fell into disrepute in the 1970s (1). Thus while *C. difficile* has assumed its proper recognition as the dominant cause of AAC, staphylococcal enterocolitis should also be considered in the diagnosis of AAC. In clinical settings where these tissue culture assays are not available, direct visualization of the colon should be relied upon to diagnose AAC.

Patients may show abnormal results of barium contrast enemas late in the course of AAC, but we believe this procedure may be hazardous because of possible perforation. The most definitive procedure to perform in patients after the onset of severe diarrhea is proctosigmoidoscopy. This technique may reveal either only a few or hundreds of friable yellowish-white plaques per colonic segment. Histological examination of a colonic mucosa biopsy is also important as early changes of AAC may exist that are inapparent at sigmoidoscopy. Colonoscopy should always be performed if the protosigmoidoscopy is normal in patients who are clinically toxic from their diarrhea. We and others have been able to identify AAC by colonoscopy when lesions were beyond the reach of the proctosigmoidoscope (25). It is most important that a definite diagnosis of AAC be established as several treatment modalities can then be more promptly employed.

Treatment

Treatment can be divided into specific and nonspecific measures. Important nonspecific measures should include discontinuation of the inducing antibiotic(s) or other drug(s), and the intravenous delivery of electrolytes and other fluids to combat dehydration. While parenteral corticosteroids were shown to be somewhat effective in the hamster model, thus far their use had not been extensively studied in humans with AAC. Agents that diminish intestinal motility such as atropine, belladonna, or diphenoxylate should also be avoided. Diarrhea is believed to be a protective reflex in this setting and should not be retarded.

Cholestyramine ameliorated diarrhea in 12 patients with AAC (22), and in the hamster it had only modest effects (7). We have used cholestyramine only in a few patients prior to recognizing that vancomycin was an effective treatment for AAC. Possible uses for cholestyramine may include treating patients with AAC who are clinically nontoxic or who have relapsed during or following oral vancomycin therapy. Colostomy, ileostomy, or even a colectomy may be required when life-threatening diarrhea occurs that is refractory to therapy or when toxic megacolon and colonic perforation(s) are present. The major treatment goal is to eradicate the fecal clostridia with antibiotics; currently vancomycin appears to be the drug of choice in treating clostridial (11, 18, 23, 25) and staphylococcal (24) AAC. Diarrhea and clinical toxicity disappear within 3–5 days of starting vancomycin therapy as does the toxin of *C. difficile* in fecal filtrates.

Virtually all isolated *C. difficile* are sensitive to vancomycin at levels less than 4 μg/ml of broth media (26). In hamsters challenged with lethal doses of antibiotics, *C. difficile* and its toxin disappear from the feces if treated once daily with oral vancomycin (21, 27). Very high concentrations of vancomycin (2000–5000 μg/ml of fecal filtrate) are achieved in the feces of humans given this drug orally (23, 24). Vancomycin is not absorbed well orally and is nontoxic by this route. The appropriate length of treatment is not well established but is between 7 and 10 days, and at least 125 to 250 mg should be given orally four times a day to adults. The minimum oral dose of vancomycin effective in AAC still needs to be established. Its only major disadvantages are a bitter aftertaste and a cost of between $40 and $80 per day. Relapses are also being reported following the use of vancomycin therapy (28). Their occurrence may be related to reinfection from the environment, to location of *C. difficile* in colonic areas where it can elude vancomycin exposure (or by sporulation?), or to host-resistance factors.

Conclusions

In the past decade *C. difficile* has been implicated in the etiology of an increasing number of enterocolitic inflammatory diseases. *C. difficile* is the major cause of AAC, which is diagnosed visually most often by proctosigmoidoscopy and/or colonoscopy. Tissue culture assay for the detection of *C. difficile* cytotoxins in fecal samples of patients with diarrhea is also supportive in the work-up and assists in treatment. Patients who are severely ill and unresponsive to simple modalities, such as hydration and the withdrawal of the inducing drug, should be treated with oral vancomycin. Prevention of this disease may lie in the development of a toxoid.

Literature Cited

1. Bartlett, J. G., Gorbach, S. L. 1977. Pseudomembranous enterocolitis (antibiotic-related colitis). *Adv. Intern. Med.* 22:455–76
2. Hall, I. C., O'Toole, E. O. 1935. Intestinal flora in newborn infants: with a description of a new pathogenic anaerobe, *Bacillus difficilis. Am. J. Dis. Child.* 49:390–402
3. Lawrence, G., Shann, F., Freestone, D. S., Walker, P. D. 1979. Prevention of necrotizing enteritis in Papua New Guinea by active immunization. *Lancet* 1:227–30
4. Small, J. D. 1968. Fatal enterocolitis in hamsters given lincomycin hydrochloride. *Lab. Animal Care* 18:411–20
5. Bartlett, J. G., Onderdonk, A. B., Cisneros, R. L., Kasper, D. L. 1977. Clindamycin-associated colitis due to a toxin-producing species of Clostridium in hamsters. *J. Inf. Dis.* 136:701–5
6. Lusk, R. H., Fekety, R., Silva, J., Browne, R. A., Ringler, D. H. Abrams, G. D. 1978. Clindamycin-induced enterocolitis in hamsters. *J. Infect. Dis.* 137:464–75
7. Fekety, R., Silva, J., Toshniwal, R., Allo, M., Armstrong, J., Browne, R., Ebright, J., Rifkin, G. 1979. Antibiotic-associated colitis: effects of antibiotics on *Clostridium difficile* and the disease in hamsters. *Rev. Infect. Dis.* 1:386–96
8. Rifkin, G. D., Silva, J., Fekety, R. 1978. Gastrointestinal and systemic toxicity of fecal extracts from hamsters with clindamycin-induced colitis. *Gastroenterology* 74:52–57
9. Allo, M., Silva, J., Fekety, R., Rifkin, G. D., Waskin, H. 1979. Prevention of clindamycin-induced colitis in hamsters by *Clostridium sordellii* antitoxin. *Am. Gastroenterol. Assoc.* 76:352–55
10. Larson, H. E., Price, A. B. 1977. Pseudomembranous colitis: presence of colostridial toxin. *Lancet* 2:1312–14
11. Rifkin, G. D., Fekety, F. R., Silva, J., Sack, R. B. 1977. Antibiotic-induced colitis. Implication of a toxin neutralized by *Clostridium sordellii* antitoxin. *Lancet* 2:1103–6
12. Taylor, N. S., Bartlett, J. G. 1979. Partial purification and characterization of a cytotoxin from *Clostridium difficile. Rev. Infect. Dis.* 1:379–385
13. Bartlett, J. G., Change, T. W., Gurwith, M., Gorbach, S. L., Onderdonk, A. B. 1978. Antibiotic-associated pseudomembranous colitis due to toxin-producing clostridia. *N. Engl. J. Med.* 298:531–34
14. LaMont, J. T., Trnka, Y. M. 1980. Therapeutic implications of *Clostridium difficile* toxin during relapse of chronic inflammatory bowel disease. *Lancet* 1:382–83
15. Bolton, R. P., Sherriff, R. J., Read, A. E. 1980. *Clostridium difficile* associated diarrhea: a role in inflammatory bowel disease? *Lancet* 1:383–84
16. Cudmore, M., Silva, J., Fekety, R. 1980. Clostridial enterocolitis produced by antineoplastic agents in hamsters and humans. *Cur. Chem. Infect. Dis.* 2:1460–61
17. Hafiz, S., McEntegart, M. G., Morton, R. S., Waitkins, S. A. 1975. *Clostridium difficile* in the urogenital tract of males and females. *Lancet* 1:420–21
18. Mogg, G. A. G., Keighley, M. R. B., Burdon, D. W., Alexander-Williams, J., Youngs, D., Johnson, M., Bentley, S., and George, R. H. 1979. Antibiotic-associated colitis—a review of 66 cases. *Br. J. Surg.* 66:738–42
19. George, W. L., Sutter, V. L., Citron, D., Finegold, S. M. 1979. Selective and differential medium for isolation of *Clostridium difficile. J. Clin. Microbiol.* 9:214–19
20. Batts, D., Silva, J., Fekety, R. 1980. Staphylococcal enterocolitis. *Cur. Chem. Infect Dis.* 2:944–45
21. Bartlett, J. G., Chang, T. W., Onderdonk, A. B. 1978. Comparison of five regimens of treatment of experimental clindamycin-associated colitis. *J. Infect. Dis.* 138:81–86

22. Kreutzer, E. W., Milligan, F. D. 1978. Treatment of antibiotic-associated pseudomembranous colitis with cholestyramine resin. *Johns Hopkins Med. J.* 143:67–72
23. Tedesco, F. R., Markam, M., Gurwith, D. C., Bartlett, J. G. 1978. Oral vancomycin for antibiotic-associated pseudomembranous colitis. *Lancet* 2:226–28
24. Khan, M. Y., Hall, W. H. 1966. Staphylococcal enterocolitis—treatment with oral vancomycin. *Ann. Intern. Med.* 65:1–8
25. Silva, J., Batts, D. H., Fekety, F. R., Plouffe, J., Rifkin, G. D. 1980. Successful treatment of clostridial colitis with oral vancomycin. *Cur. Chem. Infect. Dis.* 2:957–58
26. Dzink, J., Bartlett, J. G. 1980. In vitro susceptibility of *Clostridium difficile* isolates from patients with antibiotic-associated diarrhea or colitis. *Antimicrob. Agents Chemother.* 17:695–98
27. Browne, R. A., Fekety, R., Silva, J., Boyd, D. I., Work, C. O., Abrams, G. 1977. The protective effect of vancomycin on clindamycin-induced colitis in hamsters. *Johns Hopkins Med. J.* 141:183–92
28. George, W. L., Volpicella, N. A., Stiner, D. B. 1979. Relapses of pseudomembranous colitis after vancomycin therapy. *N. Engl. J. Med.* 301:414–15

Ann. Rev. Med. 1981. 32:335–40
Copyright © 1981 by Annual Reviews Inc. All rights reserved

DIAGNOSIS AND TREATMENT OF HERPES SIMPLEX ENCEPHALITIS[1]

❖7402

R. Whitley, M.D.

Division of Virology, Department of Pediatrics, the University of Alabama in Birmingham, School of Medicine, Birmingham, Alabama 35294

Herpes simplex virus infections of man have been recognized since ancient Greek times; however, encephalitis caused by this virus was only first described in the early 1940s by Smith, who isolated this pathogen from infected brain tissue (1). Subsequently, herpes simplex virus infections of the central nervous system have been considered the most common cause of sporadic fatal encephalitis (2, 3). Over the past four decades numerous case reports of herpes simplex encephalitis have appeared in literature attempting to define the course of infection as well as the outcome following various modalities of experimental treatment. This overview considers recent advances in our understanding of this disease and its treatment as based primarily upon data collected over the past decade through the National Institute of Allergy and Infectious Diseases Collaborative Antiviral Study Group.

DIAGNOSIS

Characteristically, herpes simplex encephalitis occurs in both sexes (male-to-female ratio of 1:1), at all ages, in all races, and throughout the year. Infection of the brain with herpes simplex virus tends to localize in the temporal and orbital-frontal lobes of the brain (4, 5) and results in hemorrhagic necrosis, which can rapidly progress over varying periods of time (a few days to a week) generally to death. The pathogenic basis for localization of herpes simplex virus to the temporal lobe remains unknown at this time.

---

[1] These studies have been supported by NO-1-AI-22532, CA-13148 and RR-O32.

0066-4219/81/0401-0335$01.00

Based upon the necrotizing nature of the infection, many of the signs and symptoms of disease can be predicted. The most common findings are personality change, disorientation, and hallucinations associated with fluctuating levels of consciousness (2, 5). Behavioral manifestations are common (5, 6). Other nonspecific findings include akinetic mutism, dysphasia, hemiparesis, fever, and altered mentation. Many patients presenting with herpes simplex encephalitis are judged to have psychiatric disorders rather than infections of the central nervous system. Focal seizures are associated with proven disease although their occurrence is not uniform. The clinical findings of the disease are not diagnostic and, therefore, other evaluations are necessary.

The diagnostic assessment of patients with a presumed focal encephalitic process focuses on cerebrospinal fluid and neurodiagnostic examinations. Examination of the cerebrospinal fluid usually demonstrates an increased number of white cells that are predominantly mononuclear in nature. However, importantly, there are a few cases where no cellular pleocytosis was evident in the cerebrospinal fluid (5–8). Invariably, red blood cells are present with as many as 500 or even 1000 being reported (5, 9). As the disease progresses, cerebrospinal fluid protein becomes elevated; yet proteinosis may not be of significant help diagnostically in that other diseases may result in similar findings. The electroencephalogram can be of great value in establishing the diagnosis of herpes simplex excephalitis, especially in the presence of periodic high voltage discharges localized to one area of the brain. The electroencephalogram is more sensitive than other neurodiagnostic tools such as the brain or computerized axial tomographic scans (8). The electroencephalographic picture may vary rapidly over a period of a few days; such changes generally do not occur with other diseases capable of mimicking herpes simplex encephalitis. The technicium brain scan and/or computerized axial tomography can be helpful in diagnosis whereby localization, edema, hemorrhage, changes in ventricular size, and midline structural shifts may appear. It has not been uncommon for patients with technicium brain scans suggestive of herpes simplex encephalitis to have unrevealing computerized axial tomography scans. (5).

Laboratory assessment of either serum or cerebrospinal fluid specimens remains to be established as useful for diagnosis. Specifically, antibody ratios of serum to cerebrospinal fluid, although reported to be useful by some (10), have not been validated in the experience of others (11). Retrospective diagnosis, by demonstrating a four-fold increase in cerebrospinal fluid antibodies, occurs in only 80–85% of patients with proven infection and may occur as frequently as 15–20% in individuals who have other diseases. As would be expected, individuals with other central nervous system diseases can excrete herpes simplex virus peripherally, which may account for the appearance of antibodies both in serum and cerebrospinal fluid (11). Newer diagnostic tests are in various stages of development; a promising one is an excreted glycoprotein assay (12).

BRAIN BIOPSY

At present, the only conclusive means of making a diagnosis is brain biopsy. In the Collaborative Antiviral Study of 184 patients undergoing brain biopsy for presumed herpes simplex encephalitis, only 105 had confirmed disease or 57% of the overall study population. Of major significance, of the remaining 79 individuals who were biopsy-negative for herpes simplex virus, diagnoses were established in 35. Thus, a diagnosis was achieved in 76% of cases studied. Diagnoses in patients with focal signs of encephalopathy not having herpes simplex encephalitis include cryptococcal infection of the brain, abscess, arteriovenous malformations, St. Louis encephalitis, Epstein-Barr virus encephalitis, subacute sclerosing panencephalitis, mumps, lymphocyctic choriomeningitis, Reye's syndrome, and tumor (13). Most of these diseases were identified through a diagnostic brain biopsy procedure. Complications (hemorrhage) from the biopsy itself were few, occurring in less than 3% of the study population.

Additional studies can be performed on brain tissue to assist the diagnostic process. These include immunofluorescence and electron microscopy as well as routine histology. Histology in many cases, but not all, shows characteristic perivascular cuffing, lymphocytic infiltration, hemorrhagic necrosis, and neuronophagia. Intranuclear inclusions are present in only about half of the cases (5). With electron microscopy or direct immunofluorescence, diagnosis can be achieved in as many as 70% of cases within three hours of the biopsy procedure itself (5).

THERAPY

Therapy of herpes simplex encephalitis with adenine arabinoside will decrease mortality and improve morbidity. Adenine arabinoside (vidarabine, ara-A) in a controlled study was shown to be effective in reducing the mortality rate from 70% in placebo recipients to 28% in drug recipients (14). These findings were statistically significant 30 days after onset of therapy ($P = 0.03$), as shown in Figure 1. Although the number of placebo recipients in this study was small, the British Working Party on herpes simplex encephalitis recently demonstrated in a double-blind controlled

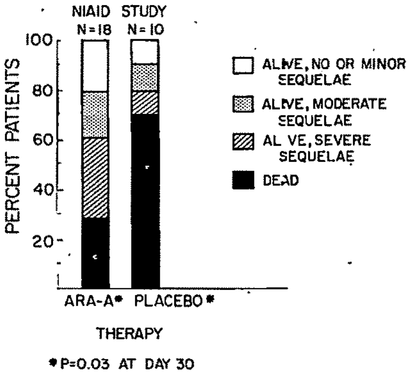


Figure 1 Morbidity and mortality in biopsy-proven herpes simplex encephalitis. (From Ref. 13, used by permission).

investigation of cytosine arabinoside that mortality both in drug and placebo recipients was in excess of 70% (15).

Data from all reported controlled investigations other than the adenine arabinoside treatment groups are summarized in Table 1. In all these therapeutic studies brain biopsy proof was the sole criterion for diagnosis. The mortality with placebo or either idoxuridine or cytosine arabinoside is in excess of 70% and normal survivors account for less than 9% of patients. These findings add credence to the reduction of mortality evident with adenine arabinoside therapy. The collaborative Antiviral Study Group has verified the observation of reduced mortality with adenine arabinoside therapy of proven disease. The mortality rate of 93 proven patients is identical to that demonstrated in the controlled investigation (14). Moreover, not only level of consciousness but also age are important determinants of outcome. Individuals lethargic at the time of the diagnosis and

Table 1 Summary of therapeutic trials employing brain biopsy proof for diagnosis of herpes simplex encephalitis

Treatment	Mortality[a] (no. dead/total)	Morbidity (no./survivors)		
		Severe	Moderate	None
Placebo (USA) (14)	7/10	1/3	1/3	1/3
Placebo (England) (15)	8/11	1/3	2/3	0/3
Cytosine Arabinoside (England) (15)	8/12	3/4	1/4	0/4
Idoxuridine (USA) (16)	5/6	1/1	0/1	0/1
	28/39 (72%)	6/11 (55%)	4/11 (36%)	1/11 (9%)

[a] One year after disease.

institution of therapy and those less than 30 years of age had the best prognosis, as only 10% die and over 60% return to a prior level of function. Lethargic individuals, irrespective of age, return to the same level of function as prior to the onset of the encephalitis in 50% of all cases studied. Delaying diagnosis and institution of therapy until semicoma or coma ensue decreases the number of normal individuals at one year of life to approximately 20% overall (13). On long-term follow-up, significant neurologic improvement is evident in all survivors of the disease process, particularly in those who are young.

Adenine arabinoside is administered intravenously at a dosage of 15 mg/kg/day for a period of ten days. The drug is given over 12 hours in standard intravenous solution. The concentration should not exceed 0.7 mg/ml[3]. Clearly, supportive care is essential in the management of the patient with this disease. Such supportive care must take into consideration that progressive cerebral edema is a consequence of the primary pathologic process. Agents to combat increased intracranial pressure must be utilized. In particular elevation of the head, mannitol and/or glycerol administration, and hyperventilation to maintain a reduced arterial P_{CO_2} are essential. The use of dexamethasone has not been substantiated as effective.

Therapy of herpes simplex encephalitis with adenine arabinoside will result in decreased mortality and improved morbidity. Nevertheless, the mortality and morbidity remain significant; therefore, newer antiviral agents are currently in clinical trials to further improve the outcome. Such agents include vidarabine 5'-monophosphate and acyclovir. It is hoped that the advent of newer agents for this disease process and the development of noninvasive diagnostic procedures will further improve the outcome of this devastating infection.

Literature Cited

1. Smith, M. G., Lennette, E. H., Reames, H. R. 1941. Isolation of virus of herpes simplex and demonstration of intranuclear inclusion in case of acute encephalitis. *Am. J. Pathol.* 17:55–68
2. Olson, L. C., Buescher, E. L., Artenstein, M. S., et al. 1967. Herpes virus infections of the human central nervous system. *N. Engl. J. Med.* 277:1271–77
3. Johnson, R. T., Olson, L. C., Buescher, E. L. 1968. Herpes simplex virus infections of the nervous system. *Arch. Neurol.* 18:260–64
4. Pierce, N. F., Portnoy, B., Leeds, N. E. 1964. Encephalitis associated with herpes simplex infection presenting as a temporal-lobe mass. *Neurology* 14: 708–13
5. Whitley, R. J., Linnemann, C., Soong, S. J. et al. 1980. Clinical assessment of patients with biopsy proven herpes simplex encephalitis. Submitted
6. Drachman, D. A., Adams, R. D. 1962. Herpes simplex and acute inclusion-body encephalitis. *Arch. Neurol.* 7: 45–63
7. Marshall, W. J. S. 1967. Herpes simplex encephalitis treated with idoxuridine and external decompression. *Lancet* 2:579–80
8. Nolan, D. C., Curruthers, M. M., Lerner, A. M. 1970. Herpes virus ho-

minis encephalitis in Michigan. *New Engl. J. Med.* 282:10–13

9. Miller, J. K., Hesser, F., Tompkins, V. N. 1966. Herpes simplex encephalitis 14. *Ann. Intern. Med.* 64:92–103
10. Levin, D. P., Lauter, C. B., Lerner, A. M. 1978. Simultaneous serum and CSF antibodies in herpes simplex virus encephalitis. *J. Am. Med. Assoc.* 240: 356–60
11. Nahmias, A. J., Whitley, R. J., Visintine, A. M., et al. 1980. Laboratory assessment of patients with biopsy proven herpes simplex encephalitis. Submitted
12. Chen, A. N., Ben-Porat, T., Whitley, R. J., Soang, S. J. Husch, M. S. et al. 1980. Herpes simplex encephalitis: vidarabine treatment and diagnostic problems. Submitted
13. Whitley, R. J., Soong, S. J., Dolin, R., et al. 1977. Adenine arabinoside therapy of biopsy proven herpes simplex encephalitis: NIAID Collaborative Antiviral Study Group. *N. Engl. J. Med.* 297:289–94
14. Longson, M. 1970. Le defie dés encephalites hérpetiqies. *Ann. Microbiol. (Inst. Pasteur)* 103A:5
15. Boston Interhospital Virus Study Group and the NIAID-Sponsored Cooperative Antiviral Clinical Study. 1975. *N. Engl. J. Med.* 292:599–603

Ann. Rev. Med. 1981. 32:341–57

PATHOGENESIS OF ACUTE BACTERIAL DIARRHEAL DISORDERS[1]

❖7403

Ralph A. Giannella, M.D.[2]

Division of Gastroenterology, University of Kentucky School of Medicine, 800 Rose Street, Lexington, Kentucky 40507

INTRODUCTION

Acute bacterial diarrheal disease is a major cause of morbidity and mortality throughout the world. In underdeveloped countries, acute bacterial diarrhea is frequently disabling, recurrent, and significantly contributes to malnutrition and death. In this setting, it has been estimated that acute diarrheal disease may be responsible for 50% of infant mortality (1,2). While the mortality and impact is considerably less in the US it remains a major cause of morbidity, hospitalization, and loss of economic resources. In fact, acute diarrheal disease is second only to the common cold in causing time lost from work (3).

This paper reviews the pathogenesis of the acute bacterial diarrheal disorders, emphasizing recent advances in our understanding rather than comprehensively covering this vast subject. Because of space limitations, the individual bacterial enteropathogens cannot be considered. This paper confines itself to the mechanisms and pathogenesis of enteric bacterial infections; selected enteropathogens are considered only for illustrative purposes to emphasize particular mechanisms. The interested reader is referred to several recent symposia that comprehensively review various aspects of acute bacterial diarrheal disease (4–7).

---

[1] The US Government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this paper.

[2] Present address: Division of Digestive Diseases, University of Cincinnati College of Medicine, 231 Bethesda Avenue, Cincinnati, Ohio 45267.

ETIOLOGY AND EPIDEMIOLOGY

Ten years ago a specific etiologic diagnosis could be made in only 20–25% of cases of acute diarrhea. Each year this figure has progressively increased so that an etiologic enteropathogen can now be identified in 50–80% of cases (8). This increase is attributable to the development of newer diagnostic techniques that permit the identification and documentation of a bacterial enteropathogen and to the identification of new etiologic agents that were formerly not thought to be involved in causing acute diarrheal disease, e.g. enterotoxigenic *Escherichia coli, Yersinia enterocolitica, Bacillus cereus, Campylobacter fetus, Vibrio parahemolyticus,* the noncholera vibrios, *Aeromonas hydrophilia,* Klebsiellae, and other coliform organisms (8–21).

The etiologic spectrum and incidence of specific bacteria in the etiology of acute diarrhea will vary with the specific population studied, the geographic locale, the age of the subjects, the time of year, and the diligence with which specific etiologic agents are sought. For example, the etiologic agents of acute diarrhea are primarily viruses when children in Boston or Taiwan are studied (22, 23); enterotoxigenic *E. coli* are the primary etiologic agents recovered in children in the Tropics, in American visitors to Mexico, or in Apache Indians (24–26); Shigellae are the major etiologic agents in populations residing in custodial institutions (27). In general, however, enterotoxigenic *E. coli, E. coli* that elaborate the classical heat-labile and heat-stable enterotoxins, are quantitatively unimportant enteropathogens for sporadically occurring acute diarrhea in either children or adults in the US. In developing countries, enterotoxigenic *E. coli* may account for 50–80% of such cases. In the US, diarrhea in children under the age of 3–4 years, especially in the winter months, is usually due to rotaviruses (10, 28).

In contrast to sporadically occurring diarrhea, when outbreaks of diarrhea in the US are considered, the following pattern emerges. [Figures given are for foodborne outbreaks in 1973 but are similar for a number of other years as well (29, 30).] A bacterial etiology could be ascertained in 66% of such outbreaks. Salmonellae accounted for 26%, *Staphylococcus aureus* 15.6%, *Clostridium perfringens* 7.1%, Shigellae 6.3%, *B. cereus* 0.8%, and *V. parahemolyticus* for 0.8% (29, 30). The most common foods involved are beef, pork and pork products, fish, poultry, and dairy products. In fact, particular etiologic agents are frequently associated with particular types of food, i.e. Salmonellae with eggs and poultry; *S. aureus* with custards, puddings, and potato salads; *C. perfringens* with poorly cooked meats and gravies; *V. parahemolyticus* with raw fish and shellfish; and *B. cereus* with rice (30, 31).

In addition to foodborne outbreaks, waterborne outbreaks of acute diarrhea have also involved thousands of subjects. These involved Salmonellae (32), enterotoxigenic *E. coli* (33), Shigellae (34), and *C. fetus* (15). They serve as a reminder of our continued vulnerability to such outbreaks in the US.

PATHOGENESIS

The establishment of an enteric infection involves complex interactions between the host and the bacterium. These interactions involve various host factors, which function to prevent bacterial colonization of the bowel, and various bacterial "virulence factors," which function to overcome host defenses.

Host Factors

Important host defense mechanisms against pathogenic bacteria include normal gastric acidity, small intestinal motility, mucosal immune systems, and the normal intestinal microflora (11, 31, 35–38). Other host factors have been described but are less well understood and less clearly established as defense mechanisms.

The ability of the stomach to secrete acid and lower the pH of gastric content below 4.0 is an important defense against ingested bacteria (38). Many bacteria are promptly killed at pH less than 4.0, e.g. Salmonellae, *E. coli, V. cholerae,* etc (39). Neutralization of gastric acidity with sodium bicarbonate markedly reduces the dose of *V. cholerae* required to infect normal volunteers (40), and Salmonellae and cholera infections occur with increased frequency in subjects who have had a gastrectomy or who have achlorhydria (38, 41).

A very important defense mechanism against ingested bacteria is normal small intestinal motility. Normal peristalsis sweeps ingested bacteria out of the small bowel into the colon. This process is probably aided by the action of secreted mucus, which helps to remove inert particles from the mucosal surface. These actions keep the small intestine sparsely populated with bacteria (35–37). When small intestinal motor function is absent or inhibited, bacteria proliferate and colonize the small intestine (35, 36), experimental enteric infection becomes easier to establish (42), enteric infection may worsen, and bacterial shedding may persist longer (43).

The mucosal immune system is another defense mechanism against pathogenic bacteria. Discussion of intestinal immunity is beyond the scope of this paper but data relevant to cholera infections are described to illustrate the point. Previous infection with *V. cholerae* confers long-lasting protection against rechallenge in volunteers (44). Oral immunization of experimental animals with cholera toxin confers immunity against subsequent challenge with either live *V. cholerae* organisms (45) or with cholera toxin (46). Intestinal immunity to cholera appears to be both antitoxic and antibacterial. Antitoxic and antibacterial antibodies have been found in the small intestinal fluid and in feces of experimental animals recovering from cholera and have been shown to originate both from the mucosal immune system and from the serum (45, 47). Intestinal antibodies prevent binding of cholera toxin to its receptor sites on the brush border of the small intestinal epithelial cell (48) and also prevent *V. cholerae* organisms from adhering to the mucosal surface of the small bowel (47, 49).

The normal resident microflora of the intestine resists intrusion of and colonization by ingested microorganisms (35–37). This resistance can be overcome by antibiotic treatment and the resistance correlates with the anaerobic population especially that of Bacteroides (50). Various factors contributing to this resistance include production of toxic metabolites, especially short-chain fatty acids; competition for nutrients, competition for space, especially on the mucosal surface; and the maintenance of a low oxidation-reduction potential (35, 37).

In spite of all these host defense factors, the prevalence of enteric infection attests to the ability of bacterial enteropathogens to overcome them and to establish disease.

Bacterial Factors

Bacterial enteropathogens have many attributes and specialized functions that promote "infectivity" and permit them to overcome host defenses. Collectively these attributes are called "virulence factors." These include the ability to (*a*) elaborate enterotoxins that alter intestinal salt and water transport mechanisms (8, 11, 51–54), (*b*) adhere to and thus colonize the mucosal surface (54–69), and (*c*) invade and penetrate the intestinal mucosa (60–63). The ability of coliform organisms to elaborate enterotoxins and to adhere to the mucosa is genetically determined by extrachromosomal elements or plasmids that are transmissible to related organisms. The ability of *V. cholerae* to elaborate enterotoxin and of Salmonellae, Shigellae, and *E. coli* to invade resides in the traditional chromosomes.

Since intestinal motility is a major host mechanism of bacterial clearance, the ability of bacteria to resist being swept away in the intestinal stream is a prime virulence factor of bacterial enteropathogens. The major mechanism that operates to overcome intestinal motility is the ability of pathogenic bacteria to adhere to the mucosal surface (37, 54–59). Another mechanism that may work in concert with the ability to adhere is chemotaxis, the ability of motile bacteria to respond to chemotactic signals released by the mucosa. These chemical signals attract bacteria from the intestinal stream to the mucosal surface. Thus far, strains of *E. coli, V. cholerae,* and *Salmonella typhimurium* have been shown to respond to this mechanism (64, 65).

Other properties that may enhance virulence and help to overcome host defenses mechanisms are (*a*) motility of bacteria (64–66), (*b*) elaboration of enzymes that may aid in the passage of bacteria through the mucus coat of the intestine to reach the cell surface and substances to aid in the invasion of the mucosal cell (55, 67, 68), (*c*) resistance to phagocytosis, and (*d*) the elaboration of substances by the bacteria that interfere with normal motor function of the intestine (69). These factors operate in various animal systems, but they currently are of uncertain importance in man and further research is needed.

The bacterial elaboration of enterotoxins, bacterial adherence, and bacterial invasion of the mucosa, are important processes and are considered in greater detail below.

ENTEROTOXINS Enterotoxins are defined operationally as substances elaborated by bacteria that provoke fluid and electrolyte secretion by the intestine. This definition encompasses two major types of substances (51). The first, the traditional enterotoxins, are exotoxins that cause diarrhea by altering intestinal electrolyte transport mechanisms without causing any significant alterations in mucosal histology, e.g. cholera toxin. The second type, cytotoxins, are exotoxins that cause alterations of mucosal histology, even to the point of causing death of epithelial cells, e.g. clostridial and staphylococcal enterotoxins. The mechanisms by which the traditional enterotoxins cause diarrhea are reviewed below. The mechanisms by which the cytotoxins cause diarrhea are uncertain. All bacterial enterotoxins thus far purified have proved to be proteins and all appear to act primarily on the small intestine (8, 11, 51–54).

Strains of bacteria that elaborate enterotoxins are *V. cholerae,* noncholera vibrios, *V. parahemolyticus, Y. enterocolitica, E. coli, Klebsiella* sp., *Enterobacter* sp., *Citrobacter* sp., *Serratia* sp., *A. hydrophilia, Shigella* sp., *Salmonella* sp., *C. perfringens, S. aureus,* and *B. cereus* (8, 11, 51–54). Other bacteria may also elaborate enterotoxins but these are less certain at present. The enterotoxins with cytotoxic properties include those of *C. perfringens, S. aureus, Sh. dysenteriae, V. parahemolyticus, A. hydrophilia, B. cereus,* and some elaborated by *E. coli* (11, 51–54). Since most is known about cholera toxin regarding its structure, action at the membrane level, mechanism of activation of adenyl cyclase, and mechanisms of intestinal secretion, this toxin is discussed in depth. This seems justified because in many ways cholera toxin is a prototype of enterotoxin structure and function (51, 53, 54, 70–74).

After ingestion, the *V. cholerae* organisms adhere to the surface of and colonize the small intestine. They then elaborate cholera toxin, which irreversibly binds to a receptor located on the brush border of the villus epithelial cell. This binding interaction stimulates adenylate cyclase and increases the intracellular concentration of cyclic AMP. The latter process results in the secretion of water and electrolytes by the small intestine. Diarrhea results because the volume of fluid secreted by the small intestine exceeds the reabsorptive capacity of the colon. All the manifestations of the natural illness can be produced by administration of cholera toxin (51, 53, 54, 74).

Cholera toxin is a protein (84,000 daltons) composed of two components called A and B. The B component, responsible for binding of the molecule to the cell receptor, is composed of five identical subunits. The A component, which is sufficient to activate adenylate cyclase in broken cell preparations, consists of two fragments, A_1 and A_2, joined by a disulfide bond. The molecule is globular and is composed of a ring of five B subunits surrounding a single A subunit (70, 71). This subunit structure is very similar to that of several other toxin peptides, in all of which portions of the toxin molecule cross cell membranes to interact with components of the cell interior, thereby inducing the cellular actions of the toxins. This subject and the common features shared by these various toxins are elegantly discussed by Gill (70). Recently published evidence suggests that Shigella enterotoxin may possess a similar subunit structure (75).

Cholera toxin binds to its receptor, the sialyl residues of GM_1 ganglioside, via its B component. After conformational changes within the cell membrane and the cholera toxin molecule, hydrophobic interactions between the A component and the cell membrane cause the A component to penetrate the membrane. The disulfide bond linking the A_1 and A_2 subunits is reduced and the A_1 peptide fragment is released (70, 71). The A_1 peptide activates adenylate cyclase, located on the inner surface of the cell membrane, by a complex enzymatic reaction involving NAD, ADP-ribose, ATP, GTP, and other factors (70, 72). These reactions, alter various epithelial electrolyte transport processes, and fluid and electrolyte secretion occurs (54, 73, 74). Specifically, cholera toxin induces intestinal secretion primarily by blocking the coupled influx (from lumen to blood) of sodium and chloride across the brush border while stimulating the secretion or outward flux (cell to lumen) of chloride and water (73, 74). These processes occur in the small intestine and diarrhea ensues (51, 54). Although various investigators have attempted to ascribe absorption and secretion of salt and water in cholera to different populations of cells within the small intestine, i.e. villus vs crypt cells, this area remains unsettled and no definitive statement can be made.

Comprehensive discussions of the various aspects of cholera toxin action can be found in several excellent reviews (54, 70–74).

In addition to the progress made in our understanding of the structure and function of cholera toxin, much has been learned about *E. coli* enterotoxins (8, 9, 51, 54, 70, 72–74). Initially, *E. coli* of particular serotyes were epidemiologically incriminated as etiologic agents of childhood diarrhea. These serotypes are called classical "enteropathogenic *E. coli*" or EPEC (9), and the mechanism by which they caused diarrhea was unknown (9). Subsequently, strains of *E. coli* that elaborate enterotoxins were described and firmly established as causes of diarrhea (8, 9). These *E. coli,* called enterotoxigenic *E. coli,* elaborate two enterotoxins, a heat-labile enterotoxin (LT) and a heat-stable enterotoxin (ST). These enterotoxigenic strains are serologically unrelated to the classical EPEC, which by and large do not elaborate either LT or ST (8, 9). LT is a large molecular weight protein that is immunochemically similar to cholera toxin and apparently induces intestinal secretion by the same adenyl-cyclase-sensitive mechanism that cholera toxin does (8, 51, 54, 70, 72–74). ST, on the other hand, is a small molecular weight peptide (74, 76, 77) that is structurally and functionally different than either *E. coli* LT or cholera toxin. ST induces secretion in both the small and large intestine (74, 78), binds to intestinal receptors distinct from those of *E. coli* LT or cholera toxin (79), and induces secretion by stimulating the guanyl cyclase-cyclic GMP system (74, 80).

There are several *E. coli* ST's, differing in structure and in host specificity (81, 82). To further the confusion, other organisms also elaborate ST's, e.g. *Klebsiella* sp., *Enterobacter* sp., *Y. enterocolitica* (52, 83). The latter ST has been shown to cause alterations in intestinal transport similar to those caused by *E. coli* ST and to activate guanyl cyclase (83). It now seems evident that the term ST, defined as an exotoxic substance elaborated by bacteria that induces intestinal secretion even after heating, encompasses a chemically heterogeneous group of enterotoxins probably elaborated by a variety of organisms.

In addition to causing diarrhea by the production of LT and various ST's other strains of *E. coli* elaborate a cytotoxin distinct from both LT and ST, and a toxin similar to shigella enterotoxin. Still other strains of *E. coli* cause diarrhea by undefined mechanisms (9, 54). Recently, the classical EPEC have been shown to release substances that induce intestinal secretion in rats. The nature of these substances is unknown. The complex subject of *E. coli* diarrhea and *E. coli* enterotoxins is succinctly reviewed by Rowe (9) and Evans (54). Further information regarding the various enterotoxins can be found in several reviews (4, 5, 31, 51–54, 60–62). Suffice it to say that many enterotoxins have been described but whether or not these play an essential role in the pathogenesis of natural illness is uncertain. This is an important area for continued research.

BACTERIAL ADHERENCE In 1971, Smith & Linggood (84) demonstrated that enterotoxigenic strains of *E. coli* must possess two attributes to cause diarrheal disease in piglets: the ability to produce enterotoxins and the ability to colonize the small intestine. Elaboration of enterotoxin alone was not sufficient to cause diarrhea. They also demonstrated that adherence of *E. coli* to the surface of the small intestinal mucosa was required for the proliferation of the organisms and the colonization of the small intestine (84). These two attributes have also been shown to be essential to the production of *E. coli* diarrhea in humans (85). The ability to adhere to the surface of the intestine appears to be important to the disease produced by *E. coli, V. cholerae,* Salmonellae, and Shigellae (55–60). Of these bacterial enteropathogens, adhesive factors have been most studied in *E. coli* and *V. cholerae* and my discussion deals with these organisms. For a more comprehensive discussion of bacterial adherence, the reader is referred to several recent reviews (37, 56, 58, 65).

The most important mechanism by which enteropathogenic bacteria adhere to the intestinal mucosa involves specific ligand-receptor binding interactions between bacterial structures called pili or fimbriae and specific receptor sites on the surface of the intestinal cell. Fimbriae are fine filamentous protein structures that occur on and protrude from the surface of the bacterium like hairs (57–59). Various kinds of fimbriae are involved in adherence to the intestinal cell. These are serologically distinct from one another, which probably reflects differing protein composition.

Initially two such fimbrial antigens were appreciated in porcine strains of *E. coli,* K88 and K99, and these antigens were considered required for *E. coli* to be pathogenic for pigs. However, strains of *E. coli* pathogenic for pigs but lacking the K88 and K99 antigens were soon found and this led to the discovery of a third adhesive-fimbrial antigen, 987p (86). A fourth, yet unnamed, fimbrial antigen, has recently been described (87). A similar situation occurred with strains of *E. coli* that cause diarrhea in man. Initially a colonization factor antigen was discovered, CFA/I, that proved to be a pilus (88) and soon thereafter a second such factor was described, CFA/II (89). Some investigators thought that either of these particular colonization factors was required for virulence of enterotoxigenic *E. coli* in man. However, it is now clear that additional antigens, distinct from CFA/I and CFA/II, occur in human strains (90–92). It is likely that the list of such adhesive factors will continue to grow. Nevertheless the particular adhesive factors now appear to be species specific, i.e. K88 and K99 to porcine strains, CFA/I and CFA/II to human strains. This specificity of fimbriae to host species is probably largely responsible for the host specificity of *E. coli* strains: human strains are not pathogenic for pigs and porcine strains are not pathogenic for man (9, 57).

As summarized by Moon (59), the evidence incriminating these fimbriae as adhesive factors binding *E. coli* to the surface of the small intestine and as required factors for the causation of diarrheal disease is as follows: (*a*) the loss or gain of fimbriae by genetic means results in the loss or gain of the ability to adhere to and to colonize the intestine; (*b*) enteric infections with enterotoxigenic *E. coli* result in the production of circulating antibodies to pili; (*c*) adhesion to the intestine both in vitro and in vivo is blocked by type-specific anti-pilus antibody; (*d*) immunization of animals with purified pilus vaccines protects against subsequet challenge with virulent organisms (93); and (*e*) diarrhea is produced in human volunteers by strains bearing such pili but not by those lacking such structures (85).

The binding of enterotoxigenic *E. coli* to the intestinal epithelial cell involves specific receptors located on the brush border of the villus cell. The receptors are specific for particular fimbriae and the same intestinal cell possesses various receptors, each of which binds to a specific type of pilus. For example, the porcine small intestinal cell has at least three, and perhaps four, distinct types of fimbrial receptors—receptors specific for K88, K99, and 987p antigen (87, 94). These receptors probably involve the oligosaccharide side-chain moieties of the membrane glycoproteins and glycolipids, since the binding of some bacterial enteropathogens to intestine can be inhibited by various sugars (58, 91, 95). However, since the binding of most strains of enteropathogenic *E. coli* do not seem to be inhibited by sugars, binding to groups other than sugars is probably also occurring. Recent evidence suggests that both hydrophobic interactions and Van der Waal's forces may be involved in the binding of *E. coli* to the intestine (96, 97).

The host receptors for enterotoxigenic *E. coli* are under genetic control (97). Two phenotypes of intestinal brush borders exist in the pig, those that bind and those that do not bind enterotoxigenic *E. coli.* The ability to bind *E. coli* is inherited in a simple Mendelian manner (97). Brush borders prepared from pigs resistant to infection by K88-bearing *E. coli* do not bind the organisms, while brush borders from small animals who developed clinical illness do indeed bind the organisms. Thus, the presence or absence of such receptors may explain in part why only certain members of a population contract diarrheal illness although all are exposed, i.e. ingested bacteria cannot adhere to and thus cannot colonize the small intestine of the host lacking such receptors (97). A variety of additional factors, both bacterial and host, affect and modulate the adherence of *E. coli* to the intestinal cell. These are reviewed by Ofek & Beachey (58).

The other bacterial enteropathogen for which intestinal binding data are available is *V. cholerae.* The binding of *V. cholerae* to the intestine is calcium dependent and inhibited by L-fucose. The available data also suggest that there may be more than one intestinal receptor for the cholera organism (95, 98).

Thus, enteropathogenic bacteria must be able to adhere to the intestinal mucosa in order to colonize and establish disease. The presence of receptors on the small intestinal villus cells probably determines host susceptibility to and the development of diarrheal disease by certain strains of *E. coli.* In addition, adherence not only permits colonization but may also facilitate the delivery of enterotoxin to the epithelium and may even enhance the ability of the organism to elaborate enterotoxin (59, 65, 84).

MUCOSAL INVASION The ability to invade the intestinal mucosa defines a second major group of bacterial enteropathogens, which includes Salmonellae, Shigellae, *Y. enterocolitica, C. fetus, V. parahemolyticus,* and some strains of *E. coli.* The most studied, from the point of view of pathogenesis, are the Salmonellae and Shigellae and this discussion deals primarily with these organisms.

The overall pathophysiology of the invasive diarrheal disorders can be divided into at least six steps: (*a*) ingestion of the organisms, (*b*) colonization of and proliferation of the organisms within the intestine, (*c*) invasion of the intestinal mucosa, (*d*) intramucosal multiplication, (*e*) stimulation of intestinal secretion, and (*f*) dissemination from the intestine (31, 53, 60–63).

Some comments about the pathophysiology of the invasive bacterial diarrheal disorders must be made before the process of mucosal invasion can be considered in detail. After ingestion, the organisms colonize and proliferate within the intestinal lumen. Unlike the enterotoxigenic organisms discussed above, where the upper small intestine is the primary site colonized, invasive bacteria primarily colonize the colon. However, in salmonellosis, the ileum is colonized as well; in shigellosis, the small intestine is transiently colonized early in the course of the disease when watery diarrhea rather than dysentery is the predominant intestinal manifestation of illness (99). The organisms then invade the intestinal mucosa, Salmonellae invade the ileal and colonic mucosa while Shigellae invade the colonic mucosa. After the organisms multiply within the intestinal cells, they spread laterally to invade other epithelial cells (Shigellae) or penetrate into the lamina propria to then disseminate from the intestine (Salmonellae and *Y. enterocolitica*). As a consequence of this invasion of the mucosa and of the intramucosal multiplication of the organisms, an acute inflammatory reaction develops and mucosal ulceration can occur. Gross ulceration of the colonic mucosa commonly occurs in shigellosis, which accounts for the dysenteric stools. Gross ulceration, however, is much less common in either *Salmonella* or *Yersinia* infections. More common to the latter are microscopic and minute ulcerations involving both the ileum and colon. As a consequence of the invasion, fluid secretion and diarrhea occur. The mechanisms of the fluid secretion and diarrhea in the invasive disorders is uncertain but does not involve "exudation" or increased mucosal permeability (60, 62), explanations frequently offered. However, since Shigellae, Salmonellae, and *Y. enterocolitica* elaborate enterotoxins, they may be responsible (31, 52, 54). Regardless of the mechanism of fluid secretion, the ability of these organisms to invade the intestinal mucosa is clearly an essential prerequisite to the establishment of disease (60, 62, 100, 101).

Invasive bacteria must adhere to cells prior to invasion. This view is supported by electron microscopic studies of the invasion of intestinal and tissue culture cells by Salmonellae and Shigellae (60, 61, 63, 102–105) which demonstrate that these organisms adhere to the cell surface prior to invasion of the cell. The bacterial surface structures and properties required for this adhesion, however, are undefined. In the case of Salmonellae, the ability to adhere to the wall of the intestine does not appear to be related to pili, to bacterial cell wall lipopolysaccharide structure (LPS), or to the specific O-sugar repeat units of the LPS coat (102–105). Furthermore, the mere association or adherence of bacteria to the mucosal surface is not sufficient for mucosal invasion. This conclusion is based on the observations that many enterotoxigenic bacteria, *E. coli, V. cholerae,* etc, as well as various nonpathogenic bacteria adhere to the mucosa without invading the epithelium.

Bacterial invasion of cells involves a series of interactions between the bacterium and the host cell. It requires the active participation of both the bacterium and the host cell, and various bacterial and host factors are required (106–108). A number of mophologic and biochemical studies of intestinal and other cells demonstrate that Salmonellae and Shigellae enter cells by endocytosis (61, 106–108). Several bacterial factors have been thought important to the ability of bacteria to invade cells, but the extensive literature concerning these factors is confusing and often contradictory. The contradictory data may be explained in part by the fact that different investigators have used different experimental designs involving different strains of bacteria and different host cells. Nevertheless, it seems highly likely that specific characteristics of the bacterial cell surface and particular metabolic activities of the invading bacteria are important to the invasion of mucosal surfaces. I believe that the data can be reconciled and summarized as follows: (*a*) In the case of *Shigella flexneri,* several attributes are required for invasiveness, including an intact and complete LPS coat and specific O-sugar repeat units (62, 109), though intact LPS and appropriate

O-repeat units alone are not sufficient for invasion (62, 109). Viability of the organism and its active metabolism are also required (107). This may involve the elaboration of a factor that signals and stimulates the target cell to initiate the endocytotic-invasive process (68). (*b*) In the case of Salmonellae, neither an intact LPS coat nor specific O-sugar repeat units are required for invasion (102, 103). However, like Shigellae, viable and metabolically active organisms are required. An extracellular product elaborated by the Salmonellae may be involved in the invasive process, since the brush border degenerates prior to any contact with the Salmonella organism (61). (*c*) In both Salmonellae and Shigellae, metabolic activity of the host cell, involving microfilaments and glycolysis, is also required for invasion (106–108).

Bacterial invasion of the mucosa cannot establish disease—the invading organisms must also multiply intracellularly. Strains of *Sh. flexneri* that can invade but cannot multiply do not cause disease when fed to the relevant host (62, 109).

After invasion and multiplication, fluid secretion ensues. The mechanism of secretion is uncertain but probably is a multifactorial process involving components of the acute inflammatory reaction (31, 60, 110, 111) and elaboration of enterotoxins (31, 51, 52, 54, 60, 99, 101). The papers by Giannella et al (60, 63, 110, 111), Formal and co-workers (62, 101), Turnbull (31), Keusch (112), and Peterson & Sandefur (113) consider the mechanisms of fluid secretion in detail. Detailed discussions of the clinical and pathophysiological features of the illnesses caused by invasive *E. coli, Y. enterocolitica, C. fetus,* and *V. parahemolyticus* can be found elsewhere (9, 14, 16–18, 83, 114, 115).

SUMMARY

Acute bacterial diarrheal disease is a worldwide problem of enormous magnitude. In recent years a number of bacteria have been added to the list of recognized etiologic agents causing acute diarrheal disease. This was made possible by our increased understanding of the mechanisms by which such bacteria cause diarrhea and by the development of methods to detect these bacterial enteropathogens. We are now able to define an etiologic agent in 50–80% of cases of acute diarrhea, depending on the particular population. The bacterial agents recently incriminated as important causes of diarrhea include *E. coli, Y. enterocolitica, B. cereus, C. fetus, V. parahemolyticus,* and many other coliform organisms.

Establishment of an enteric infection depends upon a complex interplay between host defense mechanisms and bacterial virulence factors adapted to overcome these defenses. Bacterial enteropathogens cause diarrhea primarily by elaborating enterotoxins (which also requires the organisms to adhere to the surface of the intestinal cell) and by invading the intestinal mucosa.

The number of known bacterial enterotoxins has rapidly increased. Enterotoxins cause intestinal secretion and diarrhea by stimulating the adenyl cyclase system or the guanyl cyclase system and by other mechanisms yet to be defined. The ability of enterotoxigenic bacteria to adhere to the intestine involves a specific binding interaction between bacterial structures called pili or fimbriae and specific receptors on the surface of intestinal cells. Both bacterial pili and the intestinal receptors are under genetic control.

A variety of other bacteria, Salmonellae, Shigellae, *Y. enterocolitica* etc, must invade the mucosa to cause diarrheal disease. The ability to invade is essential to the pathogenesis of disease and requires particular surface characteristics of the bacterium as well as the active participation of both the bacterium and the host cell. The bacteria probably elaborate substances that signal the host cell to initiate the invasive process, i.e. endocytosis. The mechanism by which invasive bacteria evoke intestinal secretion is uncertain but is probably a multifactorial process involving products elaborated by the mucosal acute inflammatory reaction and enterotoxins elaborated by the bacteria.

Literature Cited

1. Gordon, J. E. 1971. Diarrheal disease of early childhood—worldwide scope of the problem. *Ann. NY Acad. Sci.* 176:9–15
2. Rohde, J. E., Northrup, R. S. 1976. Taking science where the diarrhea is. See Ref. 4, pp. 339–66
3. Dingle, J. A., Badger, G. F., Jordan, W. S. 1964. *Illness in the Home: A study of 25,000 illnesses in a group of Cleveland families*, pp. 19–32; 188–218. Cleveland: Western Reserve Univ. Press. 398 pp.
4. Elliott, K., Knight, J., eds. 1976. *Acute Diarrhea in Childhood*, Ciba Found. Symp. 42. Amsterdam: Elsevier. 375 pp.
5. Lambert, H. P., ed. 1979. *Infections of the GI Tract.* Clinics in Gastroenterology 8:3, pp. 549–838. Philadelphia: Saunders
6. Janowitz, H. D., Sachar, D. B. 1979. *Frontiers of Knowledge in the Diarrheal Diseases.* Upper Montclair, NJ: Projects in Health, Inc. 383 pp.
7. Field, M., Fordtran, J. S., Schultz, S. D., eds. 1980. *Secretory Diarrhea.* Bethesda, MD: Am. Physiol. Soc. 227 pp.
8. Sack, R. B. 1975. Human diarrheal disease caused by enterotoxigenic *E. coli.* *Ann. Rev. Microbiol.* 29:333–53
9. Rowe, B. 1979. The Role of *E. coli* in gastroenteritis. See Ref. 5, pp. 625–44
10. Banatvala, J. E. 1979. The role of viruses in acute diarrheal disease. See Ref. 5, pp. 569–98
11. Plotkin, G. R., Kluge, R. M., Waldman, R. H. 1979. Gastroenteritis: etiology, pathophysiology, and clinical manifestations. *Medicine* 58:95–114
12. DuPont, H. L. 1979. Etiologic diagnosis of acute diarrhea. *Ann. Intern. Med.* 88:707–8
13. Pai, C. H., Mors, V. 1978. Production of enterotoxin by Yersinia enterocolitica. *Infect. Immun.* 19:908–911
14. Vantrappen, G., Agg, H. O., Ponette, E., Geboes, K., Bertrand, P. 1977. Yersinia enteritis and enterocolitis: gastroenterological aspects. *Gastroenterology* 72:220–27
15. Haley, C. E., Gunn, R. A., Hughes, J. M. Lippy, E. C., Craun, G. F. 1980. Outbreaks of waterborne disease in the U.S., 1978. *J. Infect. Dis.* 141:794–97
16. Carpenter, C. C. J. 1979. More pathogenic vibrios. *N. Engl. J. Med.* 300:39–41
17. Butzler, J. P., Skirrow, M. B. 1979. Campylobacter enteritis. See Ref. 5, pp. 737–65

18. Barker, W. H., Gangarosa, E. J. 1974. Food poisoning due to *Vibrio parahemolyticus. Ann. Rev. Med.* 25: 75–81
19. Hughes, J. M., Hollis, D. G., Gangarosa, E. J., Weaver, R. E. 1978. Noncholera vibrio infections in the US: clinical, epidemiologic, and laboratory features. *Ann. Intern. Med.* 88:602–6
20. Ljungh, A., Popoff, M., Wadstrom, T. 1977. Aeromonas hydrophilia in acute diarrheal disease: detection of enterotoxin and biotyping of strains. *J. Clin. Microbiol.* 6:96–100
21. Wadstrom, T., Aust-Kettis, A., Habte, D., Meecuwisse, G., Holmgren, J., Mollby, R., Soderlindo, R. 1976. Enterotoxin-producing bacteria and parasites in stools of Ethiopian children with diarrheal disease. *Arch. Dis. Child.* 51:865–70
22. Echeverria, P., Blacklow, N. R., Smith, D. H. 1975. Role of heat-labile toxigenic *E. coli* and reovirus agents in diarrhea in Boston children. *Lancet* 2:1113–16
23. Echeverria, P., Ho, M. T., Blacklow, N. R. 1977. Relative importance of viruses and bacteria in the etiology of pediatric diarrhea in Taiwan. *J. Infect. Dis.* 137:383–90
24. Guerrant, R. L., Moore, R. A., Kirschenfeld, P. M., Sande, M. A. 1975. Role of toxigenic and invasive bacteria in acute diarrhea of childhood. *N. Engl. J. Med.* 293:567–73
25. Gorbach, S. L., Kean, B. H., Evans, D. G., Evans, D. J., Bessudo, D. 1975. Traveler's diarrhea and toxigenic *E. coli. N. Engl. J. Med.* 292:933–36
26. Sack, R. B., Hirschhorn, N., Brownlee, I., Cash, R. A., Woodward, W. E., Sack, D. A. 1975. Enterotoxigenic *E. coli*-associated diarrheal disease in Apache children. *N. Engl. J. Med.* 292:1041–45
27. DuPont, H. L., Gangarosa, E. J., Reller, L. B., Woodward, W. E., Armstrong, R. W., Hammond, J., Glaser, K., Morris, G. K. 1970. Shigellosis in custodial institutions. *Am. J. Epidemiol.* 92:172–79
28. Kapikian, A. Z., Kim, H. W., Wyatt, R. G., Cline, W. L., Arrobio, J. O., Brandt, C. D., Rodriguez, W. J., Sack, D. A., Chanock, R. M., Parrott, R. H. 1976. Human-reovirus-like agent as the major pathogen associated with "winter" gastroenteritis in hospitalized infants and young children. *N. Engl. J. Med.* 294:966–72
29. Black, R. E., Cox, R. C., Horwitz, M. A. 1978. Outbreaks of foodborne disease in the U.S., 1975. *J. Infect. Dis.* 137:213–18
30. Horwitz, M. A. 1977. Specific diagnosis foodborne disease. *Gastroenterology* 73:375–81
31. Turnbull, P. C. B. 1979. Food-poisoning with special reference to Salmonella —its epidemiology, pathogenesis, and control. See Ref. 5, pp. 663–714
32. Collaborative Report. 1971. A waterborne epidemic of Salmonellosis in Riverside, California, 1965. *Am. J. Epidemiol.* 93:33–48
33. Rosenberg, M. L., Koplan, J. P., Wachsmuth, I. K., Wells, J. G., Gangarosa, E. J., Guerrant, R. L., Sack, D. A. 1977. Epidemic diarrhea at Crater Lake from enterotoxigenic *E. coli.* A large waterborne outbreak. *Ann. Intern. Med.* 86:714–18
34. Merson, M. H., Tenney, J. H., Meyers, J. D., Wood, B. T., Wells, J. G., Rymzo, W., Cline, B., DeWitt, W. E., Skaliy, P., Mallison, G. F. 1975. Shigellosis at sea: an outbreak aboard a passenger cruise ship. *Am. J. Epidemiol.* 101:165–75
35. Donaldson, R. M. Jr. 1968. Role of indigenous enteric bacteria in intestinal function and disease. In *Handbook of Physiology,* Sect. 6, Alimentary Canal, pp. 2807–37. Washington DC: Am. Physiol. Soc.
36. Gorbach, S. L. 1971. Intestinal microflora. *Gastroenterology* 60:1110–29
37. Savage, D.C. 1977. Microbial ecology of the GI tract. *Ann. Rev. Microbiol.* 31:107–33
38. Giannella, R. A., Broitman, S. A., Zamcheck, N. 1973. Influence of gastric acidity on bacterial and parasitic enteric infections. *Ann. Intern. Med.* 78:271–76
39. Giannella, R. A., Broitman, S. A., Zamcheck, N. 1972. The gastric barrier to microorganisms in man: in vivo and in vitro studies. *Gut* 13:251–56
40. Hornick, R. B., Music, S. I., Wenzel, R., Cash, R., Libonati, J. P., Snyder, M. J., Woodward, T. E. 1971. The Broad Street pump revisited: response of volunteers to ingested cholera vibrios. *Bull. NY Acad. Med.* 47:1181–91
41. Nalin, D. R., Levine, R. J., Levine, M. M., Hoover, D., Bergquist, E., McLaughlin, P., Libonati, J., Alam, J., Hornick, R. B. 1978. Cholera, non-vibrio cholera, and stomach acid. *Lancet* 2:856–59
42. Formal, S. B., Abrams, G. D., Schneider, H., Sprinz, H. 1963. Experimental shigella infections. VI. Role of

the small intestine in an experimental infection of guinea pigs. *J. Bacteriol.* 85:119–25

43. DuPont, H. L., Hornick, R. B. 1973. Adverse effect of lomotil therapy in shigellosis. *J. Am. Med. Assoc.* 226: 1525–28
44. Cash, R. A., Music, S. I., Libonati, J. P., Craig, J. P., Pierce, N. F., Hornick, R. B. 1974. Response of man to infection with vibrio cholerae. II. Protection from illness afforded by previous disease and vaccine. *J. Infect. Dis.* 130:325–33
45. Pierce, N. F., Cray, W. C., Engel, P. F. 1980. Antitoxic immunity to cholera in dogs immunized with cholera toxin. *Infect. Immun.* 27:632–37
46. Holmgren, J., Svennerholm, A. M., Ouchterlony, O., Anderson, A., Wallerstrom, G., Westerberg-Berndtsson, U. 1975. Antitoxic immunity in experimental cholera: protection and serum and local antibody responses in rabbits after enteral and parenteral immunization. *Infect. Immun.* 12:1331–40
47. Fubara, E. S., Freter, R. 1973. Protection against enteric bacterial infection by secretory IgA antibodies. *J. Immunol.* 111:395–403
48. Wu, A. L., Walker, A. W. 1976. Immunological control mechanism against cholera toxin: interference with toxin binding to intestinal receptors. *Infect. Immun.* 14:1034–42
49. Schrank, G. D., Verwey, W. F. 1976. Distribution of cholera organisms in experimental vibrio cholerae infections: proposed mechanisms of pathogenesis and antibacterial immunity. *Infect. Immun.* 13:195–203
50. Bohnhoff, M., Miller, C. P., Martin, W. R. 1964. Resistance of the mouse's intestinal tract to experimental Salmonella infection. *J. Exp. Med.* 120: 817–28
51. Banwell, J. G., Sherr, H. 1973. Effect of bacterial enterotoxins on the GI tract. *Gastroenterology* 56:467–97
52. Thorne, G. M., Gorbach, S. L. 1979. New bacterial enterotoxins and human diarrheal diseases. See Ref. 6, pp. 165–76
53. Grady, G. F., Keusch, G. T. 1971. Pathogenesis of bacterial diarrheas. *N. Engl. J. Med.* 285:831–41, 891–900
54. Evans, N. 1979. Pathogenic mechanisms in bacterial diarrhea. See Ref. 5, pp. 599–623
55. Jones, G. W. 1975. Adhesive properties of enteropathogenic bacteria. In *Microbiology 1975,* ed. D. Schlessinger, pp. 137–42. Washington, DC: Am. Soc. Microbiol. 521 pp.
56. Smith, H. 1977. Microbial surfaces in relation to pathogenicity. *Bacteriol. Rev.* 41:475–500
57. Moon, H. W., Isaacson, R. E., Pohlenz, J. 1979. Mechanisms of association of enteropathogenic *E. coli* with intestinal epithelium. *Am. J. Clin. Nutr.* 32: 119–27
58. Ofek, I., Beachey, E. H. 1980. Bacterial adherence. *Adv. Intern. Med.* 25: 503–32
59. Moon, H. W. 1980. Luminal and mucosal factors of small intestine affecting pathogenic colonization. See Ref. 7, pp. 127–39
60. Giannella, R. A. 1975. Pathogenesis of Salmonella enteritis and diarrhea. See Ref. 55, pp. 170–73
61. Takeuchi, A. 1971. Penetration of the intestinal epithelium by various microorganisms. *Curr. Top. Pathol.* 54: 1–27
62. Formal, S. B., Gemski, P., Giannella, R. A., Takeuchi, A. 1975. Studies on the pathogenesis of enteric infections caused by invasive bacteria. See Ref. 4, pp. 27–43
63. Giannella, R. A., Formal, S. B., Dammin, G. J., Collins, H. 1973. Pathogenesis of Salmonellosis: studies of fluid secretion, mucosal invasion and morphologic reaction in the rabbit ileum. *J. Clin. Invest.* 52:441–53
64. Allweiss, B., Dostal, J., Carey, K., Edwards, T., Freter, R. 1977. The role of chemotaxis in the ecology of bacterial pathogens in mucosal surfaces. *Nature* 266:448–50
65. Freter, R., O'Brien, P. C. M., Macsai, M. S. 1979. Effect of chemotaxis on the interaction of cholera vibrios with intestinal mucosa. *Am. J. Clin. Nutr.* 32: 128–32
66. Guentzel, M. D., Berry, L. J. 1975. Motility as a virulence factor for *Vibrio cholerae. Infect. Immun.* 11:890–97
67. Ross, C. 1959. Mucinase activity of intestinal organisms. *J. Pathol. Bacteriol.* 77:642–44
68. Osada, Y., Ogawa, H. 1977. Phagocytosis stimulation by an extracellular product of virulent Shigella flexneri 2a. *Microbiol. Immunol.* 21:49–55
69. Mathias, J. R., Martin, J. L., Carlson, G. M., Nogueira, J., Giannella, R. A. 1980. Repetitive bursts of action potentials in response to *E. coli* heat-stable toxin: a defense mechanism of the bacterium. *Clin. Res.* 28:281 (Abstr.)

70. Gill, D. M. 1978. Seven toxic peptides that cross cell membranes. In *Bacterial Toxins and Cell Membranes,* ed. J. Jeljaszewicz, T. Wadstrom, pp. 291–332. New York: Academic. 430 pp.
71. Fishman, P. H. 1980. Mechanism of action of cholera toxin: events on the cell surface. See Ref. 7, pp. 85–106
72. Moss, J., Vaughan, M. 1980. Mechanism of activation of adenylate cyclase by choleragen and *E. coli* heat-labile enterotoxin. See Ref. 7, pp. 107–26
73. Field, M. 1974. Intestinal secretion. *Gastroenterology* 66:1063–84
74. Field, M. 1979. Mechanisms of action of cholera and *E. coli* enterotoxins. *Am. J. Clin. Nutr.* 32:189–96
75. Olsnes, S., Eiklid, K. 1980. Isolation and characterization of *Shigella shigae* cytotoxin. *J. Biol. Chem.* 255:284–89
76. Staples, S. J., Asher, S. E., Giannella, R. A. 1980. Purification and characterization of heat-stable enterotoxin produced by a strain of *E. coli* pathogenic for man. *J. Biol. Chem.* 255:4716–21
77. Alderete, J. F., Robertson, D. C. 1978. Purification and chemical characterization of the heat-stable enterotoxin produced by porcine strains of enterotoxigenic *E. coli. Infect. Immun.* 19: 1021–30
78. Eade, M. N., Giannella, R. A. 1978. Effect of purified *E. coli* heat—stable enterotoxin on intestinal transport and possible mechanism of action. *Clin. Res.* 74:1121 (Abstr.)
79. Giannella, R. A., Luttrell, M., Drake, K. 1980. Binding of pure *E. coli* heat-stable enterotoxin to isolated rat intestinal villus cells. *Clin. Res.* 28:764 (Abstr.)
80. Giannella, R. A., Drake, K. W. 1979. Effect of purified *E. coli* heat-stable enterotoxin on intestinal cyclic nucleotide metabolism and fluid secretion. *Infect. Immun.* 24:19–23
81. Burgess, M. N., Bywater, R. J., Cowley, C. M., Mullan, N. A., Newsome, P. M. 1978. Biological evaluation of a methanol-soluble, heat-stable *E. coli* enterotoxin in infant mice, pigs, rabbits, and calves. *Infect. Immun.* 21:526–31
82. Kapitany, R. A., Forsyth, G. W., Scoot, A., McKenzie, S. F., Worthington, R. W. 1979. Isolation and partial characterization of two different heat-stable enterotoxins produced by bovine and porcine strains of enterotoxigenic *E. coli. Infect. Immun.* 26:173–77
83. Rao, M. C., Guandalini, S., Laird, W. J., Field, M. 1979. Effects of heat-stable enterotoxin of *Yersinia enterocolitica* on ion transport and cyclic guanosine 3', 5'-monophosphate metabolism in rabbit ileum. *Infect. Immun.* 26:875–78
84. Smith, H. W., Linggood, M. A. 1971. Observations on the pathogenic properties of the K88, Hly and Ent plasmids of *E. coli* with particular reference to porcine diarrhea. *J. Med. Microbiol.* 4: 467–85
85. Satterwhite, T. K., Evans, D. G., DuPont, H. L., Evans, D. J. 1978. Role of *E. coli* colonization factor antigen in acute diarrhea. *Lancet* 2:181–84
86. Nagy, B., Moon, H. W., Isaacson, R. E. 1977. Colonization of porcine intestine by enterotoxigenic *E. coli:* Selection of piliated forms in vivo, adhesion of piliated forms to epithelial cells in vitro, and incidence of a pilus antigen among porcine enteropathogenic *E. coli. Infect. Immun.* 16:344–52
87. Moon, H. W., Kohler, E. M., Schneider, R. A., Whipp, S. C. 1980. Prevalence of pilus antigens, enterotoxin types, and enteropathogenicity among K88-negative enterotoxigenic *E. coli* from neonatal pigs. *Infect. Immun.* 27:222–30
88. Evans, D. G., Silver, R. P., Evans, D. J., Chase, D. G., Gorbach, S. L. 1975. Plasmid-controlled colonization factor associated with virulence in *E. coli* enterotoxigenic for human. *Infect. Immun.* 12:656–67
89. Evans, D. G., Evans, D. J. 1978. New surface-associated heat-labile colonization factor antigen (CFA/II) produced by enterotoxigenic *E. coli* of sero groups 06 and 08. *Infect. Immun.* 21:638–47
90. Orskov, I., Orskov, F. 1977. Special O:K:H serotypes among enterotoxigenic *E. coli.* strains from diarrhea in adults and children. *Med. Microbiol. Immunol.* 163:99–110
91. Guerrant, R. L., Bergman, M. J. 1979. Attachment factors among enterotoxigenic *E. coli.* See Ref. 6, pp. 137–47
92. Levine, M. M., Rennels, M. D., Daya, V., Hughes, T. P. 1980. Hemagglutination and colonization factors in enterotoxigenic and enteropathogenic *E. coli* that cause diarrhea. *J. Infect. Dis.* 141:733–37
93. Nagy, B., Moon, H. W., Isaacson, R. E., To, C. C., Brinton, C. C. 1978. Immunization of suckling pigs against enteric enterotoxigenic *E. coli* infection by vaccination dams with purified pili. *Infect. Immun.* 21:269–74
94. Isaacson, R. E., Fusco, P. C., Brinton, C. C., Moon, H. W. 1978. In vitro adhe-

sion of *E. coli* to porcine small intestinal epithelial cells: pili as adhesive factors. *Infect. Immun.* 21:392–97

95. Jones, G. W. 1977. The attachment of bacteria to the surfaces of animal cells. In *Microbial Interactions*, ed. J. L. Reissig, Vol. 3, pp. 141–76. London: Chapman Hall
96. Smyth, C. J., Jonsson, P., Olsson, E., Soderlind, O., Rosengren, J., Hjerten, S., Wadstrom, T. 1978. Differences in hydrophobic surface characteristics of porcine enteropathogenic *E. coli* with or without K88 antigen as revealed by hydrophobic interaction chromatography. *Infect. Immun.* 22:462–72
97. Sellwood, R., Kearns, M. 1979. Inherited resistance to *E. coli* diarrhea in pigs. See Ref. 6, pp. 113–22
98. Freter, R., Jones, G. W. 1976. Adhesive properties of *Vibrio cholerae:* nature of the interaction with intact mucosal surfaces. *Infect. Immun.* 14:246–56
99. Rout, W. R., Formal, S. B., Giannella, R. A., Dammin, G. J. 1975. Pathophysiology of shigella diarrhea in the rhesus monkey: intestinal transport, morphological, and bacteriological studies. *Gastroenterology* 68:270–78
100. Labrec, E. H., Schneider, H., Magnani, T. J. Formal, S. B. 1964. Epithelial cell penetration as an essential step in the pathogenesis of bacillary dysentery. *J. Bacteriol.* 88:1503–18
101. O'Brien, A. D., Gentry, M. K., Thompson, M. R., Doctor, B. P., Gemski, P., Formal, S. B. 1979. Shigellosis and *E. coli* diarrhea: relative importance of invasive and toxigenic mechanisms. *Am. J. Clin. Nutr.* 32:229–33
102. Giannella, R. A., Washington, O., Gemski, P., Formal, S. B. 1973. Invasion of HeLa cells by *Salmonella typhimurium:* a model for study of invasiveness of Salmonella. *J. Infect. Dis.* 128:69–75
103. Kihlstrom, E., Edebo, L. 1976. Association of viable and inactivated *Salmonella typhimurium* 395 MS and MR 10 with HeLa cells. *Infect. Immun.* 14:851–57
104. Tannock, G. W., Blumershine, R. V. H., Savage, D.C 1975. Association of *Salmonella typhimurium* with, and its invasion of, the ileal mucosa of mice. *Infect. Immun.* 11:365–70
105. Hohman, A. W., Schmidt, G., Rowley, D. 1978. Intestinal colonization and virulence of Salmonella in mice. *Infect. Immun.* 22:763–70
106. Kihlstrom, E., Nilsson, L. 1977. Endocytosis of *Salmonella typhimurium* 395 MS and MR10 by HeLa cells. *Acta Pathol. Microbiol. Scand.* 85:322–28
107. Hale, T. L., Bonventre, P. F. 1979. Shigella infection of Henle intestinal epithelial cells: role of the bacterium. *Infect. Immun.* 24:879–86
108. Hale, T. L., Morris, R. E., Bonventre, P. F. 1979. Shigella infection of Henle intestinal epithelial cells: Role of the host cell. *Infect. Immun.* 24:887–94
109. Formal, S. B., Hornick, S. B. 1978. Invasive *Escherichia coli. J. Infect. Dis.* 137:641–44
110. Giannella, R. A., Gots, R. E., Charney, A. M., Greenough, W. B., Formal, S. B. 1975. Pathogenesis of Salmonella-mediated intestinal fluid secretion: activation of adenyl cyclase and inhibition by indomethacin. *Gastroenterology* 69: 1238–45
111. Giannella, R. A. 1979. The importance of the intestinal inflammatory reaction in Salmonella-mediated intestinal secretion. *Infect. Immun.* 23:140–45
112. Keusch, G. T. 1979. Shigella infections. See Ref. 5, pp. 645–62
113. Peterson, J. W., Sandefur, P. D. 1979. Evidence of a role for permeability factors in the pathogenesis of Salmonellosis. *Am. J. Clin. Nutr.* 32:197–209
114. Carter, P. B. 1975. Pathogenicity of *Yersinia enterocolitica* for mice. *Infect. Immun.* 11:164–70
115. Pai, C. H., Mors, V., Seemayer, T. A. 1980. Experimental *Yersinia enterocolitica* enteritis in rabbits. *Infect. Immun.* 28:238–44

Ann. Rev. Med. 1981. 32:359–74
Copyright © 1981 by Annual Reviews Inc. All rights reserved

PROSTAGLANDINS AND WATER EXCRETION

◆7404

Randall Mark Zusman, M.D.

Cardiac Unit, Medical Services, Massachusetts General Hospital, and
Department of Medicine, Harvard Medical School, Boston, Massachusetts
02114

Introduction

In healthy vertebrates the ionic composition of the extracellular fluid (ECF) remains within a closely regulated range of "normal" values. In response to changes in electrolyte or fluid intake, or in response to perturbations in homeostasis due to environmental factors, the kidney reabsorbs or excretes sodium, potassium, other ions, and water; in this way it maintains physiologic ionic concentrations in the ECF compartment and the total ECF volume. Vasopressin, the hormone that regulates water excretion, increases the water permeability of the luminal membrane of the renal collecting tubule (1); it is released from the pituitary in response to increased ECF osmolatity (2). Adenosine 3',5'-monophosphate (cyclic AMP) and theophylline can mimic the effects of vasopressin on the water permeability of the isolated toad urinary bladder (3). This suggests that vasopressin increases the water permeability of the epithelial cell by activating adenylate cyclase and increasing the cellular content of cyclic AMP. Handler et al (4) confirmed that both vasopressin and theophylline increase the concentration of cyclic AMP in the epithelial cells of the toad urinary bladder, and thus provided critical evidence substantiating the hypothesis that vasopressin's action is mediated by the "second messenger," cyclic AMP (1, 5, 6).

In 1965, Orloff et al (7) reported that prostaglandin (PG) E_1 inhibited the water permeability response of the toad urinary bladder to vasopressin and theophylline but had no effect on the response to cyclic AMP itself. Because the water-permeability response to vasopressin and theophylline is dependent upon adenylate cyclase activity but the response to cyclic AMP is independent of this enzyme, it was concluded that PGE_1 inhibits both basal

0066-4219/81/0401-0359$01.00

and vasopressin-stimulated adenylate cyclase activity. Further studies showed that PGE_2 also inhibited vasopressin- and theophylline-stimulated water permeability (8) and vasopressin-stimulated cyclic AMP accumulation in the epithelial cells of the isolated toad urinary bladder (9, 10). Based upon these observations and similar studies on the isolated rabbit cortical collecting tubule, Grantham & Orloff (11) suggested that endogenous prostaglandins might modulate the water-permeability response to vasopressin in vivo. The finding that the inhibition of endogenous prostaglandin synthesis increases the water flow response to vasopressin in vitro (12, 13) supported this suggestion.

Vasopressin-Stimulated Prostaglandin Biosynthesis

Although the biochemical pathways leading to the synthesis of prostaglandins E_2 and $F_{2\alpha}$, thromboxane A_2, prostacyclin, and leukotrienes have been well described (14), little is known about the mechanisms controlling the cellular regulation of prostaglandin biosynthesis. Within the kidney a minimum of five cell types (the endothelium of the arterial and venous vasculature, the mesangial cells of the glomerulus, the epithelial cells of Bowman's capsule, the renomedullary interstitial cells, and the epithelial cells of the collecting duct) can synthesize prostaglandins (15). Studies of prostaglandin biosynthesis by the intact kidney in vitro or in vivo or by renal slices in vitro have not ascertained the direct stimuli that increase prostaglandin biosynthesis. With the isolation of renomedullary interstitial cells in tissue culture (16), a model system for the study of cellular prostaglandin biosynthesis became available.

Renomedullary interstitial cells exist in high density within the inner medulla and papilla of the mammalian kidney (16, 17) and are known to contain large quantities of arachidonic acid, the biosynthetic precursor of PGE_2 (18). When isolated in tissue culture these cells synthesize and release large quantities of PGE_2 (16, 19). This synthesis is increased by the polypeptide hormones arginine vasopressin, bradykinin, and angiotensin II, and by hyperosmolality or removal of potassium from the incubation medium. Adrenal steroid hormones and nonsteroidal anti-inflammatory drugs inhibit PGE_2 synthesis by these cells (19) (Figure 1). Polypeptide hormone-stimulated PGE_2 synthesis depends on the activation of a hormone-sensitive phospholipase that catalyzes the release of arachidonic acid from the cellular arachidonic acid storage pool. Potassium, hyperosmolality, and adrenal steroid hormones affect arachidonic acid release and PGE_2 synthesis by an unknown mechanism. Whereas the effects of potassium, hyperosmolality, and adrenal steroid hormones on PGE_2 synthesis are unaltered by cycloheximide, protein synthesis inhibitors decrease peptide hormone-stimulated PGE_2 synthesis. These cellular responses suggest that peptide hormones

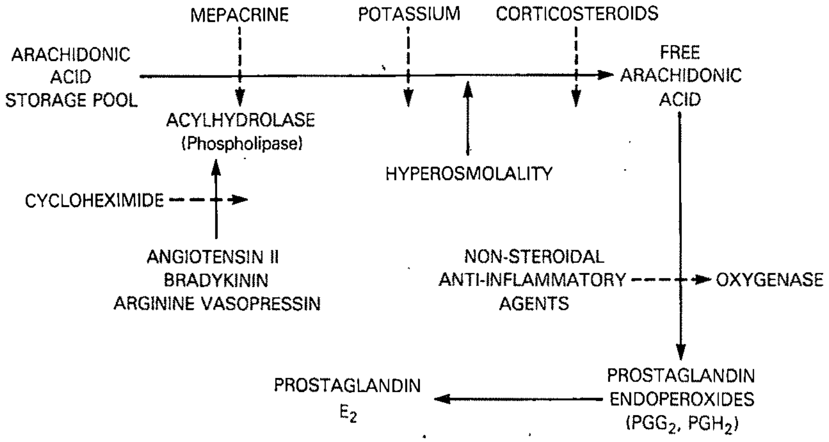


Figure 1 The regulation of PGE_2 biosynthesis by rabbit renomedullary interstitial cells in tissue culture (see text). (Reprinted from Ref. 22 with the permission of the publisher.)

interact with a cellular receptor and increase protein synthesis and phospholipase activity. On the other hand, the effects of potassium and hyperosmolality are probably due to changes in the membrane properties of the cells, which alter the rate of release of arachidonic acid from membrane phospholipids (19–23). The pattern of effects of vasopressin on cellular phospholipase activity, arachidonic acid release, and PGE_2 synthesis is shown in Table 1. The selective effects of mepacrine (a phospholipase inhibitor), naproxen (a cyclooxygenase inhibitor), and cycloheximide (a protein synthesis inhibitor) on cellular PGE_2 synthesis, also shown in Table 1, confirm the pathways diagrammed in Figure 1.

The finding that vasopressin stimulates PGE_2 synthesis by renomedullary interstitial cells in tissue culture, suggested that it might also stimulate PGE_2 synthesis by other cell types, including perhaps the renal collecting tubule. Using the isolated toad urinary bladder in vitro, Zusman et al (24) investigated the possibility that vasopressin stimulates PGE_2 synthesis by water-permeable membranes and that the PGE_2 synthesized modulates vasopressin-stimulated water permeability. They found that although vasopressin, theophylline, and cyclic AMP each stimulated water flow, only vasopressin stimulated PGE_2 synthesis. Furthermore, inhibition of vasopressin-stimulated PGE_2 synthesis by mepacrine increased vasopressin-stimulated water flow. The addition of exogenous arachidonic acid to the incubation medium greatly increased PGE_2 synthesis and inhibited vasopressin-stimulated water flow. Exogenous PGE_2 also inhibited vasopressin- as well as theophylline-stimulated water flow but had no effect on cyclic AMP-stimulated water flow. Studies in which the bladders were incubated with radiolabeled arachidonic acid confirmed that vasopressin

Table 1 The effects of vasopressin and protein synthesis inhibition, phospholipase inhibition or prostaglandin synthesis inhibition on prostaglandin E_2 synthesis by rabbit renomedullary interstitial cells in tissue culture[a]

Experimental conditions	Prostaglandin E_2 synthesis (picomoles PGE_2/μg protein/hr)	Phospholipase activity (femtomoles arachidonic acid released/μg protein/hr)	Radiolabeled materials released (femtomoles/μg protein/hr)	
			Arachidonic acid	PGE_2
Basal	1.1 ± 0.3	355 ± 25	52 ± 6	8.4 ± 0.4
Arginine vasopressin (0.5 μM)	4.0 ± 0.5*	1,028 ± 76*	725 ± 32*	117 ± 12*
+ cycloheximide (35 μM)	1.8 ± 0.2*+	584 ± 50*+	357 ± 22*+	57 ± 7*+
+ mepacrine (100 μM)	2.1 ± 0.3*+	512 ± 34*+	322 ± 46*+	51 ± 8*+
+ naproxen (100 μM)	0.1 ± 0.05*+	1,096 ± 112*	749 ± 41*	0.8 ± 0.2*+

[a] Rabbit renomedullary interstitial cells were incubated for one hour in Kreb's bicarbonate buffer (pH 8.1, potassium concentration = 3 mEq/1, osmolality=300 mOsm/l). PGE_2 synthesis was measured by radiommunoassay (19); phospholipase activity was measured as the rate of release of arachidonic acid from a radiolabeled arachidonic acid containing phospholipid substrate (21); and arachidonic acid and PGE_2 release were measured by thin-layer chromatographic analysis of the incubation medium from cells incubated with radiolabeled arachidonic acid (20). The cells were preincubated with cycloheximide (protein synthesis inhibition), mepacrine (phospholipase inhibition) or naproxen (prostaglandin synthesis inhibition) for a minimum of two hours prior to incubation with arginine vasopressin. Each value represents the mean ± standard error of the mean. $N = 12$. An asterisk (*) indicates statistical significance ($P < 0.01$) compared to basal values; a superscript + indicates statistical significance ($P < 0.01$) compared to incubation with vasopressin alone.

stimulated PGE_2 synthesis by increasing the rate of release of arachidonic acid from the cellular storage pool, presumably via activation of a cellular phospholipase (24).

In view of the ability of PGE_2 to inhibit vasopressin- and theophylline-stimulated water flow, and the effects of prostaglandin E_2 synthesis inhibitors (Figure 2), phospholipase inhibitors, or arachidonic acid on vasopressin-, theophylline- and cyclic AMP-stimulated water flow, the pathway described in Figure 3 was proposed. This schematic representation describes a counter-regulatory system in which vasopressin not only stimulates adenylate cyclase activity and cyclic AMP accumulation, thereby resulting in increased water permeability, but also stimulates arachidonic acid release and PGE_2 synthesis, which result in decreased adenylate cyclase activity and an inhibition of vasopressin-stimulated water flow. This model conforms to the experimentally derived relationships between vasopressin, theophylline, cyclic AMP, PGE_2, and prostaglandin synthesis inhibition. It describes a system in which any action that increases PGE_2 synthesis inhibits vasopressin- and theophylline-stimulated water flow but has no effect on cyclic AMP-stimulated water flow. Prostaglandin synthesis inhibition however, increases vasopressin- and theophylline-stimulated wa-

	Prostaglandin E_2	Inhibition of Prostaglandin Biosynthesis
Arginine vasopressin	↓	↑
Theophylline	↓	↑
Cyclic amp	↔	↔

Figure 2 The effects of PGE_2 and prostaglandin synthesis inhibition on vasopressin-, theophylline-, and cyclic AMP-stimulated water flow. The pattern of responses to exogenous PGE_2 or to the inhibition of prostaglandin synthesis is consistent with the inhibition by endogenous PGE_2 of basal and vasopressin-stimulated adenylate cyclase activity.

ter flow but has no effect on cyclic AMP-stimulated water flow. This model predicts that both vasopressin-stimulated and basal adenylate cyclase activity are inhibited by PGE_2 in these tissues. The inhibition of vasopressin-stimulated adenylate cyclase activity by PGE_2 has been demonstrated in renal homogenates from the hamster (25), rat (26), and rabbit (27), but it is difficult to demonstrate inhibition of basal adenylate cyclase activity by PGE_2. Recent work demonstrating that arachidonic acid itself and the endoperoxide PGH_2 inhibit basal renomedullary adenylate cyclase activity might explain in part the results of previous studies (28).

The strongest evidence for the importance of PGE_2 as a regulator of renal water reabsorption in vivo comes from studies that measured vasopressin-stimulated urinary PGE_2 excretion, and from studies on the effects of prostaglandin synthesis inhibition on vasopressin-stimulated changes in urinary osmolality and urine volume. The increase in urinary osmolality in response to vasopressin is enhanced by prostaglandin synthesis inhibition in rats (29, 30), dogs (31, 32), and humans (30, 33), and this enhanced reabsorption of water is associated with an increased accumulation of cyclic AMP within the renal medulla after prostaglandin synthesis inhibition (29). Additional evidence for the stimulation of renal prostaglandin biosynthesis by vasopressin has been obtained in rats (34, 35) and rabbits (36). In addition, the urinary excretion of PGE_2 in the vasopressin-deficient Brattleboro rat is markedly decreased compared with normal rats, but vasopressin or its analogue 1-desamino-8-D-arginine vasopressin (dD-AVP) markedly stimulated urinary PGE_2 excretion (37, 38). Moreover, prostaglandin synthesis inhibition enhances dD-AVP or vasopressin-stimulated water reabsorption in the rat; these data confirm the importance of vasopres-

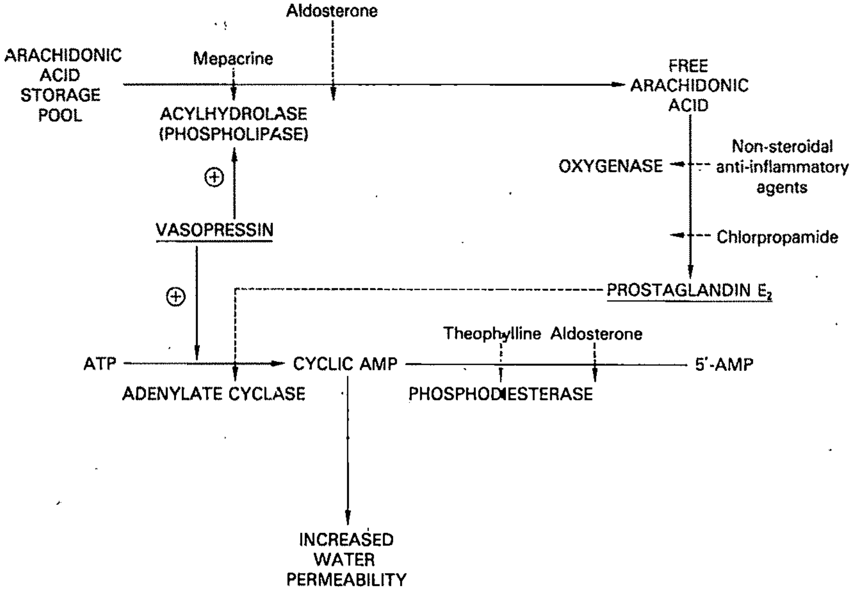


Figure 3 The role of prostaglandin E_2 in the regulation of vasopressin-stimulated water flow (see text). Agents such as mepacrine (a phospholipase inhibitor), the nonsteroidal anti-inflammatory agents (cyclo-oxygenase inhibitors), adrenal steroid hormones, or chlorpropamide decrease PGE_2 synthesis and increase vasopressin-stimulated water flow. Aldosterone also enhances vasopressin-stimulated water flow because of its inhibitory effect on cyclic nucleotide phosphodiesterase.

sin-stimulated prostaglandin biosynthesis in modulating vasopressin-stimulated adenylate cyclase activation, cyclic AMP accumulation, and water-permeability response (39).

Recent studies show that vasopressin, but not dD-AVP, stimulates PGE_2 synthesis by rat renomedullary interstitial cells in tissue culture (40). Yet dD-AVP as well as vasopressin stimulates renal PGE_2 excretion in the rat in vivo (37–39). These data imply that the increase in urinary PGE_2 excretion in response to vasopressin or dD-AVP is due to increased PGE_2 synthesis not by the renomedullary cells but rather by the vasopressin-sensitive collecting tubule. These data also suggest that PGE_2 plays an important role in modulating the response of the collecting tubule to vasopressin.

Vasopressin also stimulates thromboxane A_2 synthesis as well as PGE_2 synthesis by the toad urinary bladder in vitro (41–43). Furthermore, the inhibition of thromboxane synthesis inhibits vasopressin-stimulated water flow. These observations, as well as the findings that thromboxane analogues directly stimulate osmotic water flow, suggest that thromboxane A_2 potentiates the effects of vasopressin on water permeability, whereas

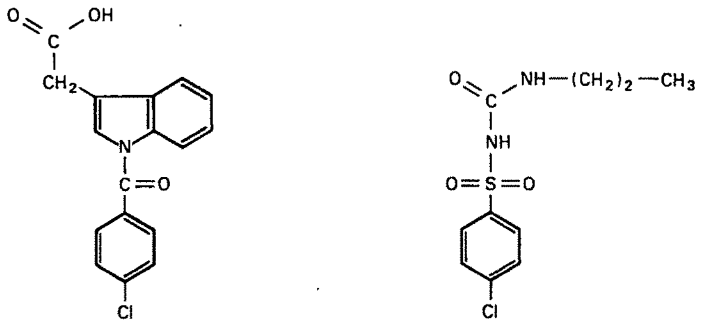
PGE_2 inhibits these effects. Nonetheless, cyclooxygenase inhibitors such as ibuprofen, indomethacin, meclofenamate, and naproxen potentiate both in vitro and in vivo vasopressin-stimulated water permeability. The overall inhibitory effects of PGE_2 thus predominate over the stimulatory or synergistic effects of thromboxane A_2 with respect to vasopressin's effect on water-permeable membranes. Whether vasopressin stimulates thromboxane A_2 synthesis by the kidney in vivo is not known; therefore the significance of thromboxane A_2 as an in vivo modulator of renal water reabsorption is unknown.

Although prostaglandins of the E series inhibit vasopressin-stimulated water permeability, infusions of PGE_1 have a paradoxical antidiuretic effect in the dog (44). This diuresis exists independently of the inhibitory effect of PGE_1 on vasopressin-stimulated adenylate cyclase activity and results from the direct stimulation of vasopressin release from the pituitary by PGE (44–46). As in the dog, intravenous infusions of PGE_2 (at doses having no hemodynamic effects) reduce urine volume and increase urine osmolality in humans (R. M. Zusman, unpublished observations). This antidiuresis is presumably due to the direct stimulation of vasopressin release from the neurohypophysis by PGE_2

Modulation of Renal Water Reabsorption in Vivo

The stimulation by vasopressin of adenylate cyclase activity and the simultaneous stimulation of the synthesis of an inhibitor of adenylate cyclase activity may appear paradoxical. But PGE_2 can be viewed as a regulator of hormone action that is itself subject to regulation by endogenous hormones and/or exogenous agents. For example, it is apparently through an effect on PGE_2 synthesis that chlorpropamide (and the other sulfonylureas) enhances the renal sensitivity to vasopressin.

Chlorpropamide is widely used to treat pituitary diabetes insipidus because it decreases urine volume (47–49). Chlorpropamide promotes the release of vasopressin from the pituitary (50) and increases the effect of vasopressin on the collecting tubule (49, 51–54). The ability of chlorpropamide to increase vasopressin- and theophylline-stimulated water flow suggested that a decrease in PGE_2 synthesis was responsible for the increase in water flow. Using the model of the toad urinary bladder, Zusman et al (55) demonstrated that chlorpropamide as well as other sulfonylureas inhibited both basal and vasopressin-stimulated PGE_2 synthesis, and that they increased theophylline- and vasopressin-stimulated water flow. Chlorpropamide inhibits the conversion of arachidonic acid to PGE_2; thus its mechanism of action resembles that of the nonsteroidal anti-inflammatory agents. The inhibition of PGE_2 synthesis by chlorpropamide accounts for the augmentation of the renal response to vasopressin (Figure 3). In view of the chemical similarities of indomethacin and chlorpropamide (Figure 4), it is



INDOMETHACIN CHLORPROPAMIDE

Figure 4 Chemical structures of indomethacin and chlorpropamide. Both indomethacin and chlorpropamide contain para-chloro-phenyl substituents and amido linkages, and both drugs are cyclooxygenase inhibitors.

not surprising that chlorpropamide and the nonsteroidal anti-inflammatory agents inhibit the conversion of arachidonic acid to PGE_2 by the same mechanism. Hence we can postulate that the ability of chlorpropamide to enhance glucose-stimulated insulin release in patients with diabetes mellitus is due to the inhibition of pancreatic prostaglandin synthesis (56), just as the ability of aspirin to increase insulin release is secondary to the inhibition of prostaglandin synthesis (57). Hyponatremia develops in patients taking sulfonylureas for the treatment of diabetes mellitus because of the direct stimulation of vasopressin release from the pituitary and the enhanced effect of vasopressin on the kidney (58).

Adrenally insufficient humans and animals are unable to excrete water loads normally. This results in reabsorption of inappropriate amounts of fluid and, as a result, hyponatremia. This chain of events is due to an inappropriately high level of circulating vasopressin (59, 60) and a fall in cardiac output that results in decreased renal blood flow (60). Although the administration of adrenal steroid hormones corrects the abnormally increased renal water reabsorption in adrenally insufficient subjects (59, 61), adrenal steroid hormones increase, rather than inhibit, vasopressin-stimulated water flow in vasopressin-sensitive epithelial membranes (62). Because adrenal steroid hormones enhance vasopressin-, theophylline-, and cyclic AMP-stimulated water flow as well as vasopressin- and theophylline-stimulated cyclic AMP accumulation in isolated epithelial cells, it was suggested that adrenal steroid hormones inhibited the activity of cyclic AMP phosphodiesterase (63). Indeed, the activity of cyclic AMP phosphodiesterase in homogenates of epithelial cells from toad urinary bladders was lower in steroid-treated tissue than in the epithelial cells of steroid-depleted tissue (64). Since the effect of steroid hormones on the phosphodiesterase was relatively small as compared with the marked enhancement of water permeability, the possibility of additional action(s) of steroid hormones was considered. Adrenal steroid hormones, which inhibit prostaglandin synthesis in vitro (19, 22), inhibit basal and vasopressin-stimulated prostaglandin biosynthesis and, therefore, increase theophylline- and vasopressin-stimulated water flow, respectively (65). Adrenal steroid hormones inhibit PGE_2 synthesis by inhibiting vasopressin-stimulated arachidonic acid release; thus, they apparently inhibit the phospholipase-mediated release of arachidonic acid. The ability of adrenal steroid hormones to enhance vasopressin-stimulated water flow even under conditions of total inhibition of prostaglandin synthesis with a nonsteroidal anti-inflammatory agent such as naproxen is due to the inhibition of phosphodiesterase by the steroid (Figure 3). The overall resistance of the adrenal steroid-deficient toad urinary bladder to the action of vasopressin results from the extraordinarily high basal and vasopressin-stimulated synthesis of PGE_2 by steroid-depleted tissue as compared to steroid-treated tissue (65) (Figure 5). Schwartz & Kokko (66) demonstrated that the cortical collecting tubule of adrenalectomized rabbits is relatively vasopressin resistant but that a response to vasopressin can be restored by aldosterone. This suggests that adrenal steroid hormones facilitate vasopressin-stimulated water permeability in vitro and in vivo via inhibition of vasopressin-stimulated arachidonic acid release and PGE_2 synthesis.

Since adrenal steroid hormones increase vasopressin-stimulated water reabsorption, one might expect that disease states in which adrenal steroid secretion is excessive would be characterized by hyponatremia. This has not been reported in the medical literature; however, the simultaneous inhibition of prostaglandin synthesis, and stimulation of adrenal steroid hormone production in sodium-depleted normal humans (67) results in a syndrome similar to the "syndrome of inappropriate antidiuretic hormone secretion" (68). Adrenocorticotrophic hormone (ACTH) and indomethacin were simultaneously administered to normal volunteers on a diet containing 2500 ml fluid, 9 meq sodium, and 100 meq potassium. The adrenal steroid hormone production increased dramatically and the subjects' urine osmolality increased markedly. Further, the subjects retained water, gained weight, and had lower plasma osmolality and sodium concentrations. Upon discontinuation of the ACTH, the steroid hormone production fell, the subjects promptly underwent a diuresis and lost weight, and their plasma osmolality and sodium concentrations were restored to normal despite continuing indomethacin administration (Figure 6). These phenomena do not occur if ACTH alone is administered to sodium-depleted normal volunteers. Although prostaglandin synthesis inhibition results in a fall in 24-

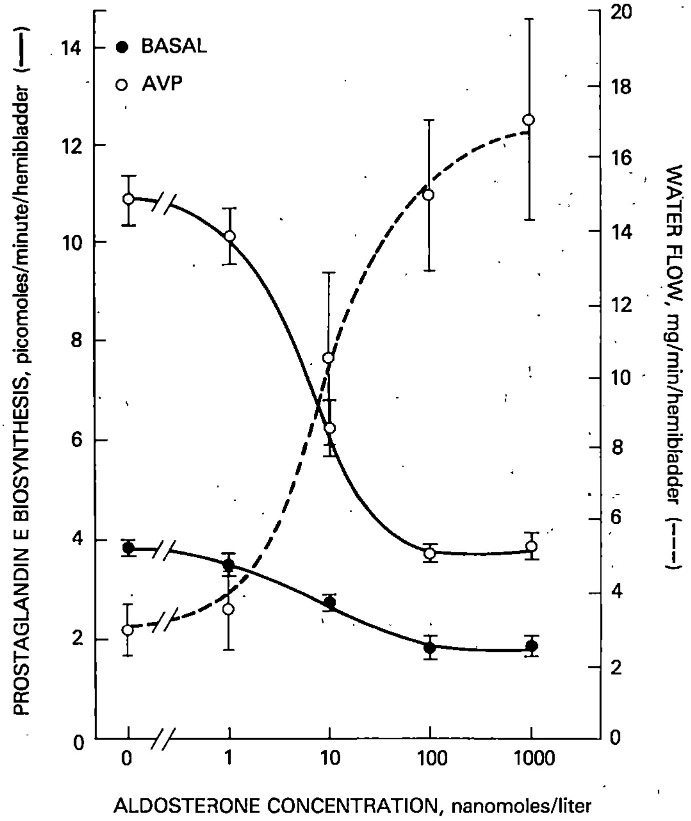


Figure 5 Effects of aldosterone basal and vasopressin (0.8 mU/ml)-stimulated PGE_2 synthesis and water flow. Isolated toad urinary bladders were incubated overnight as previously described (65); the water-permeability response and increase in PGE_2 synthesis were determined after incubation with arginine vasopressin (0.8 mU/ml).

hour urinary vasopressin excretion (R. M. Zusman and G. Robertson, unpublished observations), the administration of ACTH to sodium-depleted normal rats, but not to vasopressin-deficient Brattleboro strain rats, increases urinary osmolality (69). Thus, the increase in urinary osmolality evoked by ACTH and presumably mediated by an increase in plasma adrenal steroid hormone levels is probably dependent on vasopressin.

Although hypokalemia is associated with an increase in renal PGE_2 synthesis, the defect in urinary concentrating ability occurring in hypokalemic nephropathy does not result from the inhibitory effect of PGE_2 on vasopressin-stimulated water permeability (65, 70, 71). Similarly, although lithium, catecholamines, probenecid, demeclocycline, and cholera toxin alter vasopressin-stimulated water permeability, they have no effect on vasopressin-stimulated PGE_2 synthesis (65). Finally, although calcium in-

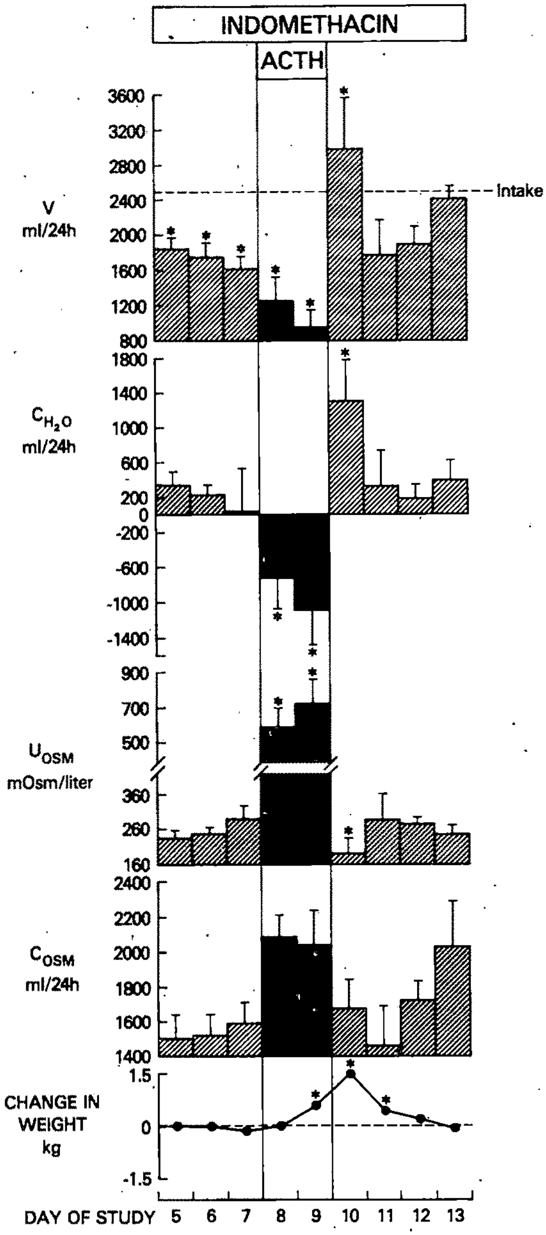


Figure 6 The effects of adrenocorticotrophic hormone and prostaglandin synthesis inhibition with indomethacin on water reabsorption in sodium-depleted normal human volunteers. The simultaneous administration of indomethacin and ACTH resulted in a statistically significant (* = $P<0.005$) fall in urine volume (V) and free water clearance (CH_2O) and a rise in urinary osmolality (Uosm) and body weight in comparison to the effects of ACTH alone. (Reprinted from Ref. 67 with permission of the publisher.)

hibits vasopressin-stimulated water flow in vitro, and hypercalcemic nephropathy in vivo is associated with vasopressin resistance, calcium does not affect vasopressin-stimulated water flow by altering the relationship between vasopressin and PGE_2 (65, 72). In contrast, the increased vasopressin-stimulated water-permeability response of isolated toad urinary bladders incubated in buffer solutions at pH 8.4, in contrast to pH 7.4, is due to the decreased PGE_2 synthesis at the higher pH (73)

Some syndromes of hereditary or acquired nephrogenic diabetes insipidus might be secondary to increased renal synthesis and the antagonism of vasopressin activity by PGE_2. Fichman et al (74) and Blachar et al (75) reported increased urinary PGE_2 excretion in patients who had hereditary nephrogenic diabetes insipidus and found a decrease in urine volume in these patients after treatment with prostaglandin synthesis inhibitors. Whether the increase in urinary PGE_2 excretion in these patients was due to a dissociation between the ability of vasopressin to stimulate adenylate cyclase activity and its ability to stimulate PGE_2 synthesis, or whether the rise was simply due to the effects of a high urinary flow rate (76, 77), is not known.

Conclusions

Prostaglandin E_2 antagonizes vasopressin-stimulated adenylate cyclase activity and water flow in isolated epithelial membranes in vitro and in the intact kidney in vivo. Paradoxically, vasopressin stimulates PGE_2 synthesis by these same tissues, thus simultaneously stimulating adenylate cyclase and the synthesis of an inhibitor of vasopressin-stimulated adenylate cyclase activity. The apparent significance of this relationship is that PGE_2 serves as a modulating hormone in vivo for the vasopressin-mediated increase in renal collecting tubule water permeability and water reabsorption. The significance of the interaction between these two substances in vivo is evident by virtue of the changes in water permeability and reabsorption after prostaglandin synthesis inhibition with nonsteroidal anti-inflammatory agents or with sulfonylurea compounds, such as chlorpropamide. Of particular note is the fact that the administration of ACTH to sodium-depleted normal subjects who had previously received indomethacin evokes an experimental model of the syndrome of inappropriate antidiuretic hormone secretion; under these conditions prostaglandins clearly contribute to the prevention of the inappropriate reabsorption of water by the kidney. On the other hand, excessive renal production of PGE_2 might account for the resistance to vasopressin exhibited by patients with acquired or hereditary nephrogenic diabetes insipidus. Although PGE_2 currently is thought to be only a minor factor in the regulation of pituitary vasopressin release, abnormalities of vasopressin release might be secondary to changes in PGE_2 synthesis by the brain.

Literature Cited

1. Handler, J. S., Orloff, J. 1973. The mechanism of action of antidiuretic hormone. In *Handbook of Physiology: Renal Physiology* (*Section 8*), ed. J. Orloff, R. Berliner, pp. 791–814. Washington, DC: Am. Physiol. Soc.
2. Schrier, R. W., Berl, T., Anderson, R. J. 1979. Osmotic and nonosmotic control of vasopressin release. *Am. J. Physiol.* 236:F321–32
3. Orloff, J., Handler, J. S. 1962. The similarity of effects of vasopressin, adenosine 3',5' phosphate (cyclic AMP) and theophylline on the toad bladder. *J. Clin. Invest.* 41:702–9
4. Handler, J. S., Butcher, R. W., Sutherland, E. W., Orloff, J. 1965. The effect of vasopressin and of theophylline on the concentration of adenosine 3',5'-phosphate in the urinary bladder of the toad. *J. Biol. Chem.* 240:4524–26
5. Orloff, J., Handler, J. S. 1964. The cellular mode of action of antidiuretic hormone. *Am. J. Med.* 36:686–97
6. Orloff, J., Handler, J. S. 1967. The role of adenosine 3',5'-phosphate in the action of antidiuretic hormone. *Am. J. Med.* 42:757–68
7. Orloff, J., Handler, J. S., Bergstrom, S. 1965. Effect of prostaglandin (PGE_1) on the permeability response of toad bladder to vasopressin, theophylline, and adenosine 3',5'-monophosphate. *Nature* 205:397–98
8. Urakabe, S., Takamitsu, Y., Shirai, D., Yuasa, S., Kimura, G., Orita, Y., Abe, H. 1975. Effect of different prostaglandins on the permeability of the toad urinary bladder. *Comp. Biochem. Physiol.* 52:1–4
9. Omachi, R. S., Robbie, D. E., Handler, J. S., Orloff, J. 1974. Effect of ADH and other agents on cyclic AMP accumulation in toad bladder epithelium. *Am. J. Physiol.* 226:1152–57
10. Lipson, L. C., Sharp, G. W. G. 1971. Effect of prostaglandin E_1 on sodium transport and osmotic water flow in the toad bladder. *Am. J. Physiol.* 220:1046–52
11. Grantham, J. J., Orloff, J. 1968. Effect of Prostaglandin E_1 on the permeability response of the isolated collecting tubule to vasopressin, adenosine 3',5'-monophosphate, and theophylline. *J. Clin. Invest.* 47:1154–61
12. Albert, W. C., Handler, J. S. 1974. Effect of PGE_1, indomethacin, and polyphloretin phosphate on toad bladder response to ADH. *Am. J. Physiol.* 226:1382–86
13. Flores, A. G. A., Sharp, G. W. G. 1972. Endogenous prostaglandins and osmotic water flow in the toad bladder. *Am. J. Physiol.* 223:1392–97
14. Samuelsson, B., Ramwell, P., Paoletti, R., eds. 1980. *Advances in Prostaglandin and Thromboxane Research.* Vols. 6, 7, 8. New York: Raven. 1816 pp.
15. Smith, W. L., Bell, T. G. 1978. Immunohistochemical localization of the prostaglandin forming cyclooxygenase in renal cortex. *Am. J. Physiol.* 235:F451–57
16. Muirhead, E. E., Germain, G., Leach, B. E., Pitcock, J. A., Stephenson, P., Brooks, B., Brosius, W. L., Daniels, E. G., Kinman, J. W. 1972. Production of renomedullary prostaglandins by renomedullary interstitial cells grown in tissue culture. *Circ. Res.* 30–31, Suppl. II:161–71
17. Nissen, H. M., Bojesen, I. 1969. On lipid droplets in renal interstitial cells. IV. Isolation and identification. *Z. Zellforsch. Mikrosk. Anat.* 97:274–84
18. Comai, K., Farber, S. J., Paulsrud, J. R. 1975. Analyses of renal medullary lipid droplets from normal, hydronephrotic, and indomethacin treated rabbits. *Lipids* 10:555–61
19. Zusman, R. M., Keiser, H. R. 1977. Prostaglandin biosynthesis by rabbit renomedullary interstitial cells in tissue culture: Stimulation by angiotensin II, bradykinin, and arginine vasopressin. *J. Clin. Invest.* 60:215–23
20. Zusman, R. M., Keiser, H. R. 1977. Prostaglandin E_2 biosynthesis by rabbit renomedullary interstitial cells in tissue culture: Mechanism of stimulation by angiotensin II, bradykinin, and arginine vasopressin. *J. Biol. Chem.* 252:2069–71
21. Zusman, R. M., Brown, C. A. 1980. The role of phospholipase in the regulation of prostaglandin E_2 biosynthesis by rabbit renomedullary interstitial cells in tissue culture: Effects of angiotensin II, potassium, hyperosmolality, dexamethasone, and protein synthesis inhibition. *Adv. Prost. Thromb. Res.* 6:243–48
22. Zusman, R. M., Keiser, H. R. 1980. Regulation of prostaglandin E_2 synthesis by angiotensin II, potassium, osmolality and dexamethasone. *Kidney Int.* 17:277–83
23. Zusman, R. M. 1980. Prostaglandin E_2 biosynthesis by renomedullary interstitial cells: In vitro studies and pathophysiologic correlations. In *The Renal Papilla and Hypertension,* ed. A. K.

Mandal, S.-O. Bohman, pp. 187–207. New York: Plenum
24. Zusman, R. M., Keiser, H. R., Handler, J. S. 1977. Vasopressin-stimulated prostaglandin E biosynthesis in the toad urinary bladder: effect on water flow. *J. Clin. Invest.* 60:1339–47
25. Marumo, F., Edelman, J. S. 1971. Effects of Ca^{++} and prostaglandin E_1 on vasopressin activation of renal adenyl cyclase. *J. Clin. Invest.* 50:1613–20
26. Beck, N. P., Kaneko, T., Zor, U., Field, J. B., Davis, B. B. 1971. Effects of vasopressin and prostaglandin E_1 on the adenyl cyclase-cyclic 3',5'-adenosine monophosphate system of the renal medulla of the rat. *J. Clin. Invest.* 50:2461–65
27. Kalisker, A., Dyer, D. C. 1972. Inhibition of the vasopressin-activated adenyl cyclase from renal medulla by prostaglandins. *Eur. J. Pharmacol.* 20:143–146
28. Herman, C. A, Zenser, T. V., Davis, B. B. 1980. Effects of prostaglandin H_2, prostaglandin E_2, and arachidonic acid on parathyroid hormone and antidiuretic hormone activation of rat kidney adenylate cyclase. *Metabolism* 29:1–8
29. Lum, G. M., Aisenbrey, G. A., Dunn, M. J., Berl, T., Schrier, R. W., McDonald, K. M. 1977. *In vivo* effect of indomethacin to potentiate the renal medullary cyclic AMP response to vasopressin. *J. Clin. Invest.* 59:8–13
30. Berl, T., Raz, A., Wald, H., Horowitz, J., Czackes, W. 1977. Prostaglandin synthesis inhibition and the action of vasopressin: Studies in man and rat. *Am. J. Physiol.* 232:F529–37
31. Anderson, R. S., Berl, T., McDonald, K. M., Schrier, R. W. 1975. Evidence for an *in vivo* antagonism between vasopressin and prostaglandin in the mammalian kidney. *J. Clin. Invest.* 56:420–26
32. Fejes-Toth, G., Magyar, A., Walter, J. 1977. Renal response to vasopressin after inhibition of prostaglandin synthesis. *Am. J. Physiol.* 232:F416–23
33. Kramer, H. J., Backer, A., Hinzen, S., Dusing, R. 1978. Effects of inhibition of prostaglandin synthesis on renal electrolyte excretion and concentrating ability in healthy man. *Prost. Med.* 1:341–49
34. Bell, C., Mya, M. K. M. 1977. Release by vasopressin of E-type prostaglandins from the rat kidney. *Clin. Sci. Mol. Med.* 52:103–6
35. Walker, L. A., Whorton, A. R., Smigel, M., France, R., Frolich, J. C. 1978. Antidiuretic hormone increases renal prostaglandin synthesis in vivo. *Am. J. Physiol.* 235:F180–85
36. Lifschitz, M. D., Stein, J. H. 1977. Antidiuretic hormone stimulates renal prostaglandin E synthesis in the rabbit. *Clin. Res.* 25:440A
37. Dunn, M. J., Greely, H. P., Valtin, H., Kinter, L. B., Beeuwkes, R. III. 1978. Renal excretion of prostaglandins E_2 and F_{2a} in diabetes insipidus rats. *Am. J. Physiol.* 235:E624–27
38. Dunn, M. J., Kinter, L. B., Shier, D., Beeuwkes, R. III. 1979. The interactions of vasopressin and renal protaglandins in the homozygous diabetes insipidus rat. *Clin. Res.* 27:496A
39. Kinter, L. B., Dunn, M. J., Shier, D., Beeuwkes, R. III. 1980. Prostaglandins, vasopressin and the renal excretion of water in the diabetes insipidus rat. *Am. J. Physiol.* In press
40. Beck, T. R., Hassid, A., Dunn, M. J. 1980. The effect of arginine vasopressin and its analogues on the synthesis of PGE_2 by rat renal medullary interstitial cells in culture. *Am. J. Physiol.* In press
41. Burch, R. M., Knapp, D. R., Halushka, P. W. 1979. Vasopressin stimulates thromboxane synthesis in the toad urinary bladder: effects of imidazole. *J. Pharmacol. Exp. Ther.* 210:344–48
42. Burch, R. M., Knapp, D. R., Halushka, P. V. 1980. Vasopressin-stimulated water flow is decreased by thromboxane synthesis inhibition or antagonism. *Am. J. Physiol.* 239:F160–67
43. Burch, R. M., Halushka, P. V. 1980. Thromboxane and stable prostaglandin endoperoxide analogs stimulate water permeability in the toad urinary bladder. *J. Clin. Invest.* In press
44. Berl, T., Schrier, R. W. 1973. Mechanism of effect of prostaglandin E_1 on renal water excretion. *J. Clin. Invest.* 52:463–71
45. Gagnon, D. J., Cousineau, D., Boucher, P. J. 1973. Release of vasopressin by angiotensin II and prostaglandin E_2 from the rat neuro-hypophysis in vitro. *Life Sci.* 12:487–97
46. Yamamoto, M., Share, L., Shade, R. E. 1976. Vasopressin release during ventriculo-cisternal perfusion with prostaglandin E_2 in the dog. *J. Endocrinol.* 71:325–31
47. Webster, B., Bain, J. 1970. Autidiuretic effect and complications of chlorpropamide therapy in diabetes insipidus. *J. Clin. Endocrinol. Metab.* 30:215–27
48. Ardvino, F., Ferraz, F. P. I., Rodriques, J. 1966. Antidiuretic action of chlor-

propamide in idiopathic diabetes insipidus. *J. Clin. Endocrinol. Metab.* 26:1325–28

49. Danisi, G., Genta, E. N., Timoner, J., Marcondes, M. 1970. Effect of chlorpropamide on urinary excretion of water and solute in patients with diabetes insipidus and on water flow across isolated toad bladder. *J. Clin. Endocrinol. Metab.* 30:528–30
50. Moses, A. M., Numann, P., Miller, M. 1973. Mechanism of chlorpropamide-induced antidiuresis in man: evidence for release of ADH and enhancement of peripheral action. *Metab. Clin. Exp.* 22:59–66
51. Hays, R. M., Ingelfinger, J. R. 1969. Evidence that prostaglandin E and vasopressin act at a chlorpropamide-sensitive site in the toad bladder. *J. Clin. Invest.* 48:37A
52. Mendoza, S. A., Brown, C. F. Jr. 1974. Effect of chlorpropamide on osmotic water flow across toad bladder and the response to vasopressin, theophylline, and cyclic AMP. *J. Clin. Endocrinol. Metab.* 38:883–89
53. Urakabe, S., Shirai, D. 1971. Effect of vasopressin, cyclic 3',5'-AMP and chlorpropamide on water permeability of toad urinary bladder. *Med. J. Osaka Univ.* 21:151–59
54. Lozada, E. S., Govanx, J., Fianki, N., Appel, G. B., Hays, R. M. 1972. Studies of the mode of action of the sulfonylureas and phenacetamides in enhancing the effect of vasopressin. *J. Clin. Endocrinol.* 34:704–12
55. Zusman, R. M., Keiser, H. R., Handler, J. S. 1977. Inhibition of vasopressin-stimulated prostaglandin E biosynthesis by chlorpropramide in the toad urinary bladder: mechanism of enhancement of vasopressin-stimulated water flow. *J. Clin. Invest.* 60:1348–53
56. Zusman, R. M., Keiser, H. R., Handler, J. S. 1977. A hypothesis for the molecular mechanism of action of chlorpropamide in the treatment of diabetes mellitus and diabetes insipidus. *Fed. Proc.* 13:2728–29
57. Robertson, R. P., Metz, S. A. 1979. Prostaglandins, the glucoreceptor, and diabetes. *N. Engl. J. Med.* 301:1446–47
58. Schrier, R. W., 1976. Water metabolism. *Kidney Int.* 10:1–132
59. Ahmed, A. B. J., George, B. C., Gonzales-Auvert, C., Dingman, J. F. 1967. Increased plasma arginine vasopressin in clinical adrenocortical insufficiency and its inhibition by glucosteroids. *J. Clin. Invest.* 46:111–23
60. Boykin, J., deTorrente, A., Erickson, A., Robertson, G., Schrier, R. W. 1978. Role of plasma vasopressin in impaired water excretion of glucosteroid deficiency. *J. Clin. Invest.* 62:738–44
61. Green, H. H., Harrington, A. R., Valtin, H. 1970. On the role of antidiuretic hormone in the inhibition of acute water diuresis in adrenal insufficiency and the effects of gluco- and mineralocorticoids in reversing the inhibition. *J. Clin. Invest.* 49:1724–36
62. Handler, J. S., Preston, A. S., Orloff, J. 1969. Effect of adrenal steroid hormones on the response of the toad's urinary bladder to vasopressin. *J. Clin. Invest.* 48:823–33
63. Stoff, J. S., Handler, J. S., Orloff, J. 1972. The effect of aldosterone on the accumulation of adenosine 3',5'-cyclic monophosphate in toad bladder epithelial cells in response to vasopressin and theophylline. *Proc. Natl. Acad. Sci. USA* 69:805–8
64. Stoff, J. S., Handler, J. S., Preston, A. S., Orloff, J. 1973. The effect of aldosterone on cyclic nucleotide phosphodiesterase activity in toad urinary bladder. *Life Sci.* 13:545–52
65. Zusman, R. M., Keiser, H. R., Handler, J. S. 1978. Effect of adrenal steroids on vasopressin-stimulated PGE synthesis and water flow. *Am. J. Physiol.* 234:F532–40
66. Schwartz, M. J., Kokko, J. P. 1979. The mechanism of the concentration defect in mineralo- and glucocorticoid deficient states. *Clin. Res.* 27:795A
67. Zusman, R. M., Vinci, J. M., Bowden, R. E., Horwitz, D., Keiser, H. R. 1979. The effect of indomethacin and adrenocorticotrophic hormone on renal function in man: An experimental model of inappropriate antidiuresis. *Kidney Int.* 15:62–70
68. Zerbe, R., Stropes, L., Robertson, G. 1980. Vasopressin function in the syndrome of inappropriate antidiuresis. *Ann. Rev. Med.* 31:315–27
69. Khosla, S., Kinter, L. B., Zusman, R. M., Beeuwkes, R. 1980. Effect of adrenocorticotrophic hormone and prostaglandins on renal water reabsorption. *Clin. Res.* 28:452A
70. Berl, T., Linas, S. L., Aisenbrey, G. A., Anderson, R. J. 1977. On the mechanism of polyuria in potassium depletion: the role of polydipsia. *J. Clin. Invest.* 60:620–25
71. Berl, T., Aisenbrey, G. A., Linas, S. L. 1980. Renal concentrating defect in the

hypokalemic rat is prostaglandin independent. *Am. J. Physiol.* 238:F37–41

72. Berl, T., Schrier, R. 1978. Calcium-prostaglandin interactions on the renal action of vasopressin: evidence for independent mechanisms. *Clin. Res.* 26: 457A
73. Forrest, J. N. Jr., Goodman, D. B. P. 1980. Prostaglandin E_2 mediates the effect of pH on ADH-stimulated water flow but ADH does not stimulate prostaglandin E_2 production in the toad urinary bladder. *Clin. Res.* 27:445A
74. Fichman, M. P., Speckart, P., Zia, P., Lee, A. 1976. Antidiuretic response to prostaglandin inhibition by indomethacin in primary and nephrogenic diabetes insipidus. *Endocrinology* 98:272
75. Blachar, Y., Zadik, Z., Shemech, M., Kaplan, B. S., Levin, S. 1980. The effect of inhibition of prostaglandin synthesis on free water and osmolar clearances in patients with hereditary nephrogenic diabetes insipidus. *Int. J. Pediatr. Nephrol.* 1:48–52
76. Lifschitz, M. D., Patak, R. V., Rosenblatt, S. G., Stein, J. H. 1978. Water diuresis stimulates renal prostaglandin E excretion in the dog. *Clin. Res.* 25:805A
77. Kirschenbaum, M., Senos, E. R. 1980. Effects of alterations in urine flow rate on prostaglandin E excretion in conscious dogs. *Am. J. Physiol.* 238:F107–11

Ann. Rev. Med. 1981. 32:375–85
Copyright © 1981 by Annual Reviews Inc. All rights reserved

PROSTAGLANDINS AND THE DUCTUS ARTERIOSUS

❖7405

P. M. Olley, M.D. and F. Coceani, M.D.

The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada

INTRODUCTION

In early embryonic life six pairs of aortic arches develop sequentially and are then modified to form the major thoracic arteries of the fetus. The distal segment of the left sixth arch persists to form the ductus arteriosus, one of the largest of fetal vessels. Widely patent until birth, the ductus then constricts so powerfully that its lumen is completely obliterated within the first 24 hours; thereafter the vessel undergoes fibrosis and in later life is represented by the ligamentum arteriosus. The behavior of the ductus in the perinatal period is of profound importance to the pathophysiology of respiratory and cardiac diseases of the newborn.

Long regarded as a passive process, fetal patency of the ductus has recently been shown to be actively maintained by prostaglandins. This discovery underlies the clinical use of prostaglandin (PG) E_2, or its structurally related PGE_1, in the newborn to preserve ductus patency, and in the use of the PG synthesis inhibitor indomethacin to induce its closure.

THE DUCTUS ARTERIOSUS IN THE FETUS

During fetal life 90% of right ventricular output passes through the ductus arteriosus to the descending aorta; the amount and direction of ductal flow is determined by the large size of the vessel, the high resistance of the fetal pulmonary circulation, and the relatively low systemic vascular resistance. The ductus is normally fully dilated in the fetus (1). Largely uninvestigated is the concept that changes in ductal tone could adjust fetal pulmonary blood flow. By shunting more than half the combined ventricular output away from the lungs the ductus reduces the total workload on the heart by a similar amount. This is probably its most important function (2).

0066-4219/81/0401-0375$01.00

The structure of the ductus arteriosus contrasts markedly with that of the adjacent aorta and pulmonary artery. The wall of the ductus is typified by the presence of abundant smooth muscle whose arrangement has been the subject of numerous studies (3, 4). Von Hayek (4) described a spiral arrangement while others described an inner circular layer and an outer, thinner, longitudinal layer. Our own studies suggest the most likely arrangement to be groups of spiral muscle of variable pitch anchored at the aortic and pulmonary ends of the vessel (5). The ductus has a well-developed internal elastic lamina, a single layer of endothelial cells, and adrenergic nerve endings in the outer layers of its wall (6, 7).

CLOSURE OF THE DUCTUS ARTERIOSUS

Functional closure of the ductus arteriosus is achieved by contraction of its smooth muscle, which both narrows and shortens the vessel.

This contraction is triggered by the rise in arterial oxygen tension that occurs with the first breath, an effect independent of the nervous system (8–10). However, a wide variety of stimuli, including mechanical and electrical stimulation, epinephrine, norepinephrine, and hemorrhage, can also induce ductus contraction (10). Kovalcik (11) studied the responsiveness of the isolated ductus arteriosus to oxygen and vasoactive agents under various conditions using isolated strips of lamb ductus or rings of guinea pig ductus. He concluded that oxygen acts directly on the ductus, possibly by releasing an intramural vasoactive agent or by stimulating a limiting enzyme reaction in the muscle cells. Both acetylcholine and norepinephrine seemed improbable candidates for this mediator role because the oxygen action was unaffected by atropine, hexamethonium, or dibenamine. The oxygen action was further studied by Fay, who found a threshold for the oxygen effect of 35 torr in tissues from mature animals and produced evidence arguing against the presence of specialized oxygen sensitive cells in either intima or adventitia (12). Oxygen responsiveness of the ductus is age dependent: it reaches a maximum at birth and declines over the next 24 hours (13).

EXPERIMENTAL DATA

The prostaglandins are oxygenated vasoactive derivatives of polyunsaturated fatty acids. In mammalian tissues, arachidonic acid is the most abundant precursor and is esterified to phospholipids in the plasma membrane and other structural components of the cell (14). A phospholipase, activated by changes in the cell environment, releases arachidonic acid, which is then metabolized by either lipoxygenase enzymes or by a cyclo-oxygenase. This latter enzyme catalyzes the formation of a PG endoperoxide, PGG_2, which is in turn converted to another endoperoxide, PGH_2. Both endoperoxides can be transformed either enzymatically or nonenzymatically to the stable PG's, PGE_2, $PGF_{2\alpha}$, and PGD_2. Alternatively, the PG endoperoxides may be converted enzymatically to either thromboxane A_2 or to prostacyclin (PGI_2). Both these products are unstable. Vascular endothelial cells are the prime source of PGI_2, while muscle cells have a greater capacity than endothelial cells to synthesize PGE_2.

Studies of the potential role of arachidonic acid metabolites in the control of ductus muscular tone led to the now generally accepted concept that fetal patency of the ductus arteriosus is actively maintained by intramural PG. This intramural mechanism may be reinforced by PG's in the fetal circulation. Several lines of evidence support this concept (15).

The ductus arteriosus contains a cyclo-oxygenase and, possibly, lipoxygenase(s) as well. In vitro experiments with isolated preparations and tissue homogenates indicate that PGI_2 is the major product of arachidonic acid metabolism, although PGE_2 and $PGF_{2\alpha}$ are also formed. Several questions remain, including (*a*) the localization of PG-synthesizing enzymes within the ductus, (*b*) the influence of neural, hormonal, and other factors including oxygen on the direction and magnitude of arachidonic acid metabolism, and (*c*) whether or not the ductus can convert 6-keto-$PGF_{1\alpha}$, the inactive metabolite of PGI_2, into 6-keto-PGE_1. This latter compound is almost as potent as PGE_2 in relaxing ductus smooth muscle. The transformation of 6-keto-$PGF_{1\alpha}$ to 6-keto-PGE_1 may occur in the liver (16).

Nonsteroidal anti-inflammatory compounds, such as indomethacin, which interfere with cyclo-oxygenase, prevent the formation of PG endoperoxides and, therefore, that of all the endoperoxide derivatives. Indomethacin constricts the fetal ductus arteriosus in vitro and in vivo in both term and immature animals, though the effect is more marked in immature animals (15). This action is fully reversed by treatment with PGE_2 or its closely related PGE_1. Indomethacin-induced ductus constriction is similar in magnitude to the normal oxygen-triggered constriction and occurs both in vivo and in vitro with doses of inhibitor that are below the range for most unspecific effects and that are also below the therapeutic range for the anti-inflammatory action. Other cyclo-oxygenase inhibitors such as aspirin and ibuprofen have been tested less extensively but also induce ductus constriction, their effectiveness correlating with their relative potency in interfering with the cyclo-oxygenase reaction.

The actions of numerous arachidonic acid metabolites have been tested on isolated strips of fetal lamb ductus in our laboratory (17). PGE_2 is the most potent ductus arteriosus relaxant agent known, only PGE_1 being comparable. Its action exceeds that of PGI_2 by several orders of magnitude and occurs in both mature and immature tissues (18). The PG endoperoxides, by conversion to PGE_2, also relax the ductus, and their action is intermediate in potency between PGE_2 and PGI_2. Reduced glutathione, which catalyzes conversion of PGH_2 to PGE_2, is also a relaxant agent confirming the presence of a functional PGH_2 to PGE_2 isomerase in ductus arteriosus tissue.

In summary an active role for prostaglandin in maintaining ductus patency during fetal life is supported by (*a*) the presence of PG biosynthetic pathways in the ductus, (*b*) the constrictor effect of PG synthetase (cyclooxygenase) inhibitors, and (*c*) the relaxant action of PGE_2 and related agents. Present evidence favors murally produced PGE_2 as the prime meditor of ductus patency. This may require reevaluation in the light of further experimental work.

Are the PG's involved in ductus closure? $PGF_{2\alpha}$ contracts both the lamb and the calf ductus, but the threshold for this action is significantly greater than that for the relaxant action of PGE_2. Starling & Elliott (19) proposed that oxygen promoted the transformation of arachidonic acid to $PGF_{2\alpha}$, which then induced constriction of the ductus. Against this hypothesis is the constrictor action of indomethacin in a high oxygen environment, which if the Starling-Elliott proposal were true should either prevent contraction or be relaxant.

Thromboxane A_2 produced in the lungs and conveyed to target sites in the ductus arteriosus by the blood stream has also been suggested as the mediator of oxygen-induced ductus constriction (1). Thromboxane A_2 synthetic capacity in the lung increases during the perinatal period (20); however, thromboxane A_2, a powerful constrictor of other vessels, has no action on the ductus in vitro (17) and is exceedingly labile in blood. An alternative hypothesis is that PG's, specifically PGE_2, become less effective postnatally (21). The demonstrated decline in PG activity occurring during the last third of gestation (22–24) could then be regarded as a priming factor for postnatal closure. Concentrations of PG's in the blood are relatively high in the fetus and fall rapidly after birth; thus their relaxant influence on the ductus is removed (25). Closure may be regarded as a withdrawal of the relaxant action of PG's thus permitting oxygen to induce constriction.

CLINICAL APPLICATIONS

Medical Closure of the Ductus

Delayed closure of the ductus arteriosus is common in premature infants although its incidence varies from center to center. This delay in closure reflects an inadequate response of ductal smooth muscle to oxygen, coupled with maximum sensitivity to relaxant PG's. A high blood concentration of circulating PG's may be important in this situation. In the presence of an open duct, left-to-right shunting develops as pulmonary vascular resistance falls. The increase in pulmonary blood flow reduces pulmonary compliance and may precipitate left heart failure with interstitial pulmonary edema, which further reduces lung compliance. In addition to the pulmonary effects the large "run off" from the systemic circulation may seriously impair both coronary perfusion, which would produce myocardial ischemia, and perfusion of vital abdominal organs. The open ductus therefore presents a major complication in the management of prematurity both with and without the respiratory distress syndrome. While the pulmonary complications may be alleviated by assisted ventilation and high inspiratory oxygen concentrations, the price of such care may be an increase in long-term complications such as bronchopulmonary dysplasia and retrolental fibroplasia.

Typically features of left-to-right shunting through the ductus arteriosus appear about 4–6 days after birth, just as the infant is recovering from the respiratory distress syndrome. Respiratory rate again increases and the peripheral pulses become bounding despite tachycardia. The precordial pulsation becomes strikingly hyperdynamic. A systolic or continuous murmur is usually audible over the upper left chest but occasionally a large ductal shunt may be entirely silent. Hepatomegaly may develop but its absence does not exclude heart failure. The chest x ray usually shows an increased heart size and increased pulmonary vascularity with or without pulmonary edema. The diagnosis of a patent ductus arteriosus with significant left-to-right shunting is usually not difficult in the premature, although it is occasionally problematic, especially in the infant already undergoing assisted ventilation. Echocardiography is extremely helpful in confirming the diagnosis and with two-dimensional techniques the open ductus can sometimes be visualized. Cardiac catheterization is not normally necessary.

Given time, the patent ductus arteriosus of the premature will usually close spontaneously, a fact used until recently to justify a conservative approach to the problem. Now the important role of the ductus in pulmonary, myocardial, and intestinal complications of the premature has been appreciated and a much more aggressive policy is adopted.

Initial treatment consists of fluid restriction (generally kept to about 120 ml/kg/day), furosemide 1–2 mg/kg once or twice per 24 hours, and controlled positive airways pressure or intermittent positive pressure ventilation. The value of digoxin is debatable because heart failure is not due to impaired muscle function but to the immaturity of the myocardium (26). Should there be no improvement within 48 hours on this regimen, closure of the ductus either surgically or pharmacologically is usually advisable.

Surgical ligation of the ductus arteriosus of the premature has been the subject of numerous reports (27). The surgical mortality is around 2% and morbidity directly attributed to the operation is uncommon. However, the possibility of using PG synthetase inhibition to induce closure of the ductus provides an attractive alternative.

Elliott et al (28) first reported constriction of the ductus arteriosus in an infant given indomethacin and this was soon confirmed (29, 30). Subsequent experience, while confirming that indomethacin can induce complete and permanent closure of the ductus in the premature, reveals a variable success rate ranging from as low as 18% to as high as 88%. However, many of the failures did show clinical or echocardiographic evidence of transient constriction of the ductus with reduced left-to-right shunting. While there are several potential explanations for this variable response rate, one of the most important factors may be the influence of gestational age on the pharmacokinetics of indomethacin and on the activity of the PG-relaxing mechanism. Vert et al (31) while confirming our finding (32) that plasma indomethacin concentrations were not significantly different between responders and nonresponders also showed that in prematures there are route-specific variations in the bioavailability of indomethacin that render them more likely to respond to intravenous rather than to oral therapy. Their data suggest that the most rational dose regimen is indomethacin 0.1–0.2 mg IV every 24 hours for three doses.

However administered, in some infants indomethacin may not achieve concentrations sufficient to inhibit prostaglandin synthesis in the ductus and other tissues. Prematures with patent ductuses have higher blood concentrations of prostaglandins than normal prematures or adults.

Postnatal factors may also influence the responsiveness to indomethacin. Continued exposure to normal or nearly normal oxygen tension for prolonged periods prior to indomethacin administration may reduce ductal reactivity. Concomitant drug therapy may also be important. Furosemide, commonly given to these infants, inhibits PG catabolism and may promote PG synthesis while theophylline a central respiratory stimulant inhibits phosphodiesterase. Finally, indomethacin may fail to close the ductus permanently because, while preventing PGE_2 formation, it may simultaneously promote the synthesis of a ductus-dilating compound via the lipoxygenase pathway. Preliminary studies suggest this could be more than a theoretical possibility (24).

The effects of indomethacin are widespread. Potential complications include renal insufficiency, bleeding diathesis, reduced formation of pulmonary surfactant, pulmonary vasoconstriction, and necrotizing enterocolitis. In practice, side-effects apart from transient oliguria appear to be uncommon.

In summary, indomethacin is a promising alternative to surgical ligation of the ductus arteriosus in premature infants. It is most likely to be effective if given intravenously at an early postnatal age. Pharmacokinetic studies suggest that it need not be given more frequently than once every 24 hours. Ideally, dosage and duration of therapy should be matched to plasma prostaglandin concentrations. The use of indomethacin in these patients should still be considered to be under investigation.

Keeping the Ductus Open

The behavior of the ductus arteriosus is critical in the pathophysiology of diverse congenital heart malformations. Easiest to understand is the type of anomaly epitomized by pulmonary atresia in which pulmonary blood flow is partially or entirely dependent on patency of the ductus. Lack of vascular continuity between the right ventricle and the pulmonary artery enforces right-to-left shunting at ventricular or atrial level, and blood can only reach the lungs via the ductus arteriosus. Similar considerations are also operative in patients with tricuspid atresia, critical pulmonary stenosis, and in complex malformations involving severe pulmonary stenosis or pulmonary atresia. These malformations can be grouped together as those having "ductus-dependent pulmonary blood flow." Such patients exhibit central cyanosis at or within a few hours of birth and thus lack the normal rise in arterial oxygen saturation—the major stimulus to ductus constriction. Despite this the ductus usually begins to close within 24–48 hours of birth, which suggests the presence of an ancillary closure mechanism. As closure begins, pulmonary blood flow falls and cyanosis increases. The resulting systemic hypoxia causes metabolic acidosis which depresses cardiac function and further reduces pulmonary blood flow by encouraging pulmonary vasoconstriction. In the absence of therapeutic intervention such infants usually die and, even if referred, are frequently extremely sick by the time they reach a specialist center. Until recently, these patients were managed by immediate cardiac catheterization followed by urgent surgical creation of some form of aortic-to-pulmonary artery shunt. Both investigation and surgery carried a high immediate mortality. Infusion of PGE_1 (identical in potency to PGE_2) to reopen the ductus and increase pulmonary blood flow now provides a new approach to the emergency care of such patients (33).

PGE_1 can be infused intravenously or intra-arterially. Side-effects are less common with the latter route while the action on the ductus appears to be identical. The usual starting dose is 0.05 μg/kg/min and the Pao_2 usually rises within minutes of starting treatment. The response is surprisingly uniform and it is usually possible to increase arterial oxygen saturation from 50% to around 80%. Pao_2 usually falls within a few minutes if the infusion is stopped. Once a response has occurred PGE_1 dosage can be reduced, often to 0.02–0.03 μg/kg/min. The improved oxygenation provides a "period of grace" during which metabolic acidosis can be reversed and arrangements for surgical palliation made on a semielective basis. Generally PGE_1 therapy needs to be maintained for 24–48 hours only, but under special circumstances prolonged administration appears to be safe and tolerance does not develop (34). More recently the use of oral PGE_2 to maintain ductus patency over several weeks or months has been reported (35). While successful in raising the Pa_{O_2}, oral therapy carries the disadvantage of frequent administration (every 1–3 hours) and the potential of vascular damage (see below).

Parenteral therapy is most likely to be effective in patients who have low initial Pa_{O_2} and who are less than four days old. Failure to respond may be due to previous complete closure of the ductus, to congenital absence or intrauterine closure of the vessel, or to extreme hypoplasia of the pulmonary arteries, which will of themselves limit pulmonary blood flow despite adequate dilatation of the ductus arteriosus.

Reported side-effects of PGE_1 include fever, bradycardia, apneic spells, hypotension, and cutaneous flushing. In the environment of a neonatal intensive care unit none of these need limit therapy and usually respond to a dose reduction.

Gittenberger-De Groot et al (36) reported histological changes in the ductus wall following PGE_1 administration. This could render the vessel more likely to rupture or undergo aneurysmal formation, but no such event has yet been reported. Pulmonary arterial smooth muscle may also be reduced by PGE_1 infusion, an effect that can be associated with the formation of localized aneurysmal dilatations (37). These histological studies require confirmation but for the present PGE_1 should only be used for as short a time as compatible with clinical needs.

PGE_1 infusions are valuable in the emergency care of newborns with other congenital heart malformations. In infants with aortic arch interruption, juxtaductal coarctation of the aorta, or hypoplastic left heart syndrome, the systemic circulation is wholly or partially ductus-dependent. Patients with transposed great arteries depend for survival on mixing between the systemic and pulmonary circulations, which may occur at atrial, ventricular, or great artery level. While the most important step in transposition is the creation of an adequate atrial septal defect, it is sometimes of benefit to preserve a patent ductus. In both these groups of patients the value of an infusion of PGE_1 has been documented. The foregoing discussion relates to the therapeutic use of PGE_1 infusions, but the compound can be of value in other situations.

Diagnostic Value of PGE_1

Severe fixed obstruction to pulmonary blood flow is common in the neonate with asplenia syndrome. Occasionally this is associated with obstructed total anomalous pulmonary venous return. Creation of a surgical systemic-to-pulmonary artery shunt in the face of such pulmonary venous obstruction may precipitate pulmonary edema and death. Neonates in whom this anatomic arrangement is suspected may be "challenged" with PGE_1. The resulting increased pulmonary blood flow precipitates pulmonary edema, which indicates that a surgical shunt is unlikely to benefit the patient and may be lethal. Accurate delineation of pulmonary arterial anatomy is crucial in any patient in whom early corrective surgery is contemplated. In patients with ductus-dependent pulmonary blood flow, angiograms taken 10–15 minutes after a PGE_1 infusion is started provide enhanced visualization of the pulmonary arteries.

Other Uses of PGE_1

PGE_1 infusions are also being evaluated as supportive therapy following certain palliative surgical procedures for patients with congenital heart lesions in which some degree of right ventricular hypoplasia is present. Infants with suspected heart disease of a type usually benefited by PGE_1 may be helped by its administration before and during transfer to a center.

SUMMARY

Fetal patency of the ductus arteriosus is an active state maintained by the relaxant action of a prostaglandin, most probably prostaglandin E_2. This PG mechanism is most active in the immature ductus and decreases toward term. The ductus closes when this prostaglandin effect is withdrawn. Indomethacin may induce closure of the patent ductus arteriosus of prematurity. It is most likely to be successful if given intravenously and early in postnatal life. Prostaglandin E_1, to maintain patency of the ductus, is now established in the emergency management of several congenital heart defects causing problems in the newborn.

Literature Cited

1. Friedman, W. F. 1978. Studies of responses of the ductus arteriosus in intact animals. In *The Ductus Arteriosus, The 75th Ross Conf. Pediatr. Res.*, ed. M. A. Heymann, A. M. Rudolph, pp. 35–43, Columbus, Ohio: Ross Laboratories
2. Heymann, M. A., Rudolph, A. M. 1975. Control of the ductus arteriosus. *Physiol. Rev.* 55:62–78
3. Sciacca, A., Condorelli, M. 1960. Involution of the ductus arteriosus. A morphological and experimental study with a critical review of the literature. *Bibl. Cardiol.* Fasc. 10
4. Von Hayek, H. 1936. Der funktionelle Ban der Nabelarterien und des Ductus Botalli. *Z. Anat. Entwicklungsgesch.* 105:15–24
5. Heaton, J. P. W. 1975. *Studies on some aspects of the structure and closure of the ductus arteriosus.* M.Sc. thesis, Univ. Toronto
6. Aronson, S., Gennser, G., Owman, C. H., Sjoberg, N. O. 1970. Innervation and contractile response of the human

ductus arteriosus. *Eur J. Pharmacol.* 11:178–86

7. Boreus, L. O., Malmfors, T., McMurphy, D. M., Olson, L. 1969. Demonstration of adrenergic receptor function and innervation in the ductus arteriosus of the human fetus. *Acta Physiol. Scand.* 77:316–21
8. Kennedy, J. A., Clark, S. L. 1941. Observations on the ductus arteriosus of the guinea pig in relation to its method of closure. *Anat. Rec.* 79:349–71
9. Kennedy, J. A., Clark, S. L. 1942. Observations on the physiological reactions of the ductus arteriosus. *Am. J. Physiol.* 136:140–47
10. Born, G. V. R., Dawes, G. S., Mott, J. C., Rennick, B. R. 1956. The constriction of the ductus arteriosus caused by oxygen and by asphyxia in newborn lambs. *J. Physiol.* 132:304–42
11. Kovalcik, V. 1963. The response of the isolated ductus arteriosus to oxygen and anoxia. *J. Physiol.* 169:185–97
12. Fay, F. S. 1971. Guinea pig ductus arteriosus I. Cellular and metabolic basis for oxygen sensitivity. *Am. J. Physiol.* 221:470–79
13. Noel, S., Cassin, S. 1976. Maturation of contractile response of ductus arteriosus to oxygen and drugs. *Am. J. Physiol.* 231:240–43
14. Moncada, S., Vane, J. R. 1979. Pharmacology and endogenous roles of prostaglandin endoperoxides, thromboxane A_2, and prostacyclin. *Pharmacol. Rev.* 30:293–330
15. Coceani, F., Olley, P. M. 1980. Role of prostaglandins, prostacyclin, and thromboxanes in the control of prenatal patency and postnatal closure of the ductus arteriosus. *Semin. Perinat.* 4:109–13
16. Wong, P. Y.-K., Malik, K. U., Sun, F. F. 1979. Hepatic metabolism of prostacyclin (PGI_2) in the rabbit: formation of a potent inhibitor of platelet aggregation. *4th Int. Prostaglandin Conf.*, Washington, DC, p. 127 (Abstr.)
17. Coceani, F., Bishai, I., White, E., Bodach, E., Olley, P. M. 1978. Action of prostaglandins, endoperoxides, and thromboxanes on the lamb ductus arteriosus. *Am. J. Physiol.* 234:H117–22
18. Coceani, F., Bodach, E., White, E., Bishai, I., Olley, P. M. 1978. Prostaglandin I_2 is less relaxant than prostaglandin E_2 on the lamb ductus arteriosus. *Prostaglandins* 15:551–56
19. Starling, M. B., Elliott, R. B. 1974. The effects of prostaglandins, prostaglandin inhibitors, and oxygen on the closure of the ductus arteriosus, pulmonary arteries and umbilical vessels in vitro. *Prostaglandins* 8:187–203
20. Powell, W. S., Solomon, S. 1978. Biosynthesis of prostaglandins and thromboxane B_2 by fetal lung homogenates. *Prostaglandins* 15:351–64
21. Coceani, F., Olley, P. M., Bodach, E. 1975. Lamb ductus arteriosus: effect of prostaglandin synthesis inhibitors on the muscle tone and the response to prostaglandin E_2. *Prostaglandins* 9: 299–308
22. Olley, P. M., White, E., Bodach, E. 1976. The contractile response of the developing lamb ductus arteriosus to ibuprofen. *Circulation* 54(2):168 (Suppl.)
23. Clyman, R. I., Mauray, F., Heymann, M. A. 1978. Ductus arteriosus: developmental response to oxygen and indomethacin. *Prostaglandins* 15:933–98
24. Coceani, F., White, E., Bodach, E., Olley, P. M. 1979. Age-dependent changes in the response of the lamb ductus arteriosus to oxygen and ibuprofen. *Can. J. Physiol. Pharmacol.* 57:825–31
25. Challis, J. R. G., Dilley, S. R., Robinson, J. S. 1976. Prostaglandins in the circulation of the fetal lamb. *Prostaglandins* 11:1041–52
26. Hoffman, J. I. E. 1977. Factors affecting shunting and the development of heart failure. See Ref. 1, pp. 69–72
27. Edmunds, L. H. Jr. 1977. Surgical management. See Ref. 1, pp. 86–89
28. Elliott, R. B., Starling, M. B., Neutze, J. M. 1975. Medical manipulation of the ductus arteriosus. *Lancet* 1:140
29. Friedman, W. F., Hirschklauw, M. J., Morton, P. P., Pitlick, P. T., Kirkpatrick, S. E. 1976. Pharmacological closure of patent ductus arteriosus in the premature infant. *N. Engl. J. Med.* 295:526–29
30. Heymann, M. A., Rudolph, A. M., Silverman, N. H. 1976. Closure of the ductus arteriosus in preterm infants by inhibition of prostaglandin synthesis. *N. Engl. J. Med.* 295:530–33
31. Vert, P., Bianchetti, G., Marchal, F., Monin, P., Morselli, P. L. 1980. Effectiveness and pharmacokinetics of indomethacin in premature newborns with patent ductus arteriosus. *Eur. J. Clin. Pharmacol.* 18:83–88
32. Alpert, B. S., Lewins, M. J., Rowland, D. W., Grant, M. J. A., Olley, P. M., Soldin, S. J., Swyer, P. R., Coceani, F., Rowe, R. D. 1979. Plasma indomethacin levels in preterm newborn infants with symptomatic patent ductus arteri-

osus—clinical and echocardiographic assessments of response. *J. Pediatr.* 95:4, 578–82

33. Olley, P. M., Coceani, F. 1980. Use of prostaglandins in cardiopulmonary diseases of the newborn. *Semin. Perinat.* 4:135–41
34. Lewis, A. B., Lurie, P. R. 1978. Prolonged prostaglandin E_1 infusion in an infant with cyanotic congenital heart disease. *Pediatrics* 61:534–36
35. Coe, J. Y., Silove, E. D. 1979. Oral prostaglandin E_2 in pulmonary atresia. *Lancet* 1:1297–98
36. Gittenberger-De Groot, A. C., Moulaert, A. J., Harinck, E. 1978. Histopathology of the ductus arteriosus after prostaglandin E_1 administration in ductus dependent cardiac anomalies. *Br. Heart J.* 40:215–20
37. Haworth, S. G., Sauer, U., Buhlmeyer, K. 1980. Effect of prostaglandin E_1 on pulmonary circulation in pulmonary atresia. A quantitative morphometric study. *Br. Heart J.* 43:306–14

Ann. Rev. Med. 1981. 32:387–404
Copyright © 1981 by Annual Reviews Inc. All rights reserved

PRIMARY RADIATION THERAPY FOR BREAST CANCER

❖7406

Jay R. Harris, M.D., Martin B. Levene, M.D., and Samuel Hellman, M.D.

Joint Center for Radiation Therapy, and Department of Radiation Therapy, Harvard Medical School, 50 Binney Street, Boston, Massachusetts 02115

Introduction

Twenty years ago the management of early carcinoma of the breast in the US was routine. Patients underwent radical mastectomy, and, usually, nothing further was done. Since then, various new concepts and approaches have been introduced, and there is now considerable controversy and uncertainty regarding the optimal treatment of this disease.

Two major areas of discussion have emerged. The first concerns the limitations of local treatment. Even when breast cancer is apparently localized, a significant proportion of patients manifest distant metastases in the years following local treatment, i.e. breast cancer is a systemic disease in many patients and no local treatment by itself can cure it. It is now clear that some form of systemic therapy, such as chemotherapy or hormonal therapy, will be required to improve the cure rate in this disease.

The second major area of discussion concerns the extent of local treatment required. In part, this reflects a recognition of the importance of body image, particularly the loss of a breast, for women in this culture. As a result of these cosmetic as well as functional concerns, surgeons have, for the most part, adopted some form of a modified radical mastectomy with preservation of one or both pectoral muscles as the standard surgical treatment for early breast cancer. There is increasing evidence that primary radiation therapy without mastectomy is an effective alternative to mastectomy as local treatment. This article reviews the historical background and current status of primary radiation therapy as local treatment for early carcinoma of the breast.

Definition of Terms

Staging refers to the grouping of patients by the extent of their disease. It is useful in determining the choice of treatment for individual patients,

0066-4219/81/0401-0387$01.00

estimating their prognosis, and comparing the results of different treatment programs. Staging of breast cancer is initially based on clinical observations, that is, the physical examination and laboratory and radiologic evaluations. The most widely used clinical staging system is the one adopted by both the UICC (International Union Against Cancer) and the AJC (American Joint Commission in Cancer Staging and End-Results Reporting). It is based on the TNM system (T, tumor; N, nodes; M, metastases) and is given in Table 1.

Relapse (or recurrence) refers to the reappearance of disease subsequent to definitive primary management. Relapse following surgery is subdivided into (*a*) local-regional, that is, on the chest wall (local) or in draining lymph node areas (regional), and (*b*) distant, for example in bone, liver, or lung.

Surgical Treatment

In assessing the value of primary radiation therapy, it is useful first to indicate the results achieved by surgery. Table 2 shows the likelihood and site of initial relapse following radical mastectomy reported in two separate series (1, 2). The results are analyzed separately for patients with histological evidence of spread to axillary nodes, and for those without spread. When nodes are not involved, local-regional relapse is uncommon (4–8%) and distant relapse is more frequent (20–22%). When nodes are involved, local-regional relapse occurs in 25–27% of patients, and distant relapse occurs in 43–58% of patients. These data emphasize three points:

Table 1 UICC-AJC staging for carcinoma of the breast[a]

Stage	Tumor	Nodes	Metastases
I	T1	N0 or N1a	M0
II	T1	N1b	M0
	T2	N0 or N1	M0
III	T1 or T2	N2	M0
	T3	N0, N1, or N2	M0
IV	any T4,	N3, or	M1

[a] T1 – tumor 2 cm or less in its greatest dimension
T2 – tumor more than 2 cm, but not more than 5 cm
T3 – tumor more than 5 cm
T4 – tumor of any size with direct extension to chest wall or skin
N0 – no palpable homolateral axillary nodes
N1 – movable homolateral axillary nodes
  a. nodes not considered to contain growth
  b. nodes considered to contain growth
N2 – fixed homolateral axillary nodes
N3 – homolateral supraclavicular or infraclavicular nodes, or edema of the arm
M0 – no evidence of distant metastases
M1 – distant metastases present.

Table 2 Likelihood and site of first relapse following radical mastectomy

	Axillary nodes		Axillary nodes	
Site of relapse	Involved[a]	Not involved[a]	Involved[b]	Not involved[b]
Local-regional only	18%	6%	25%	4%
Distant only	49%	20%	43%	20%
Local-regional and distant	9%	2%	—	—

[a] From Reference 2.
[b] From Reference 1.

1. The prognosis following surgery is highly dependent on the presence or absence of spread to axillary lymph nodes.
2. The predominant site of relapse following radical mastectomy is distant, indicating the systemic nature of the disease and the importance of developing effective systemic therapy.
3. Initial local-regional relapse following radical mastectomy is not uncommon, with an overall rate for all patients of 10–15%.

As noted above, there has been a shift in this country to the use of a modified radical mastectomy instead of the classic Halsted radical mastectomy. The long-term results achieved by the modified radical mastectomy are less well documented than those achieved by radical mastectomy. In a recent report from Johns Hopkins University, Baker et al (3) presented a retrospective comparison of the results of radical mastectomy to those of the modified radical mastectomy. For patients with early breast cancer, the local-regional relapse rate and survival at 5 years were similar for the two operations. This agrees with results obtained at other institutions, and provides the rationale for the shift toward the more conservative procedure.

Postoperative Radiation

One of the first major uses of radiation therapy for breast cancer was as an ancillary therapy following mastectomy. This use of postoperative radiation was based on two considerations. First, there is a significant incidence of local-regional relapse following mastectomy, especially if axillary nodes are involved. Second, radical mastectomy fails to remove internal mammary nodes, one of the major lymph node drainage areas from the breast. Handley (4) reported on 1000 patients with operable breast cancer undergoing internal mammary node biopsy, and found that 223 showed metastatic spread (4). Taken together these results indicate that occult cancer frequently remains in the local-regional area following radical mastectomy. It was hoped that postoperative radiation could eliminate this occult residual cancer and thereby improve the survival in those patients without distant spread.

Extensive experience with postoperative radiation clearly demonstrates that moderate-dose radiation can eliminate occult residual deposits of breast cancer. Without postoperative radiation, 20–25% of patients with positive axillary nodes will relapse in the supraclavicular area. Fletcher (5), at the M. D. Anderson Hospital, demonstrated that low-dose radiation (3000–3500 rad in 4 weeks) reduced supraclavicular relapse in axillary node-positive patients to 7%; moderate-dose radiation (4500–5000 rad in 4 weeks) reduced the relapse rate to 1%. These data stress the importance of adequate dose (4500–5000 rad in 5 weeks) in eradicating subclinical disease. The results from our own institution (6) also demonstrate the effectiveness of postoperative radiation. From July 1968 through December 1972, 354 patients were referred for postoperative radiation at the Joint Center for Radiation Therapy. These patients were referred for a variety of reasons, including spread to axillary lymph nodes, inner or central primary cancers, or because the surgeon was concerned about the margins of resection. The overall axillary, supraclavicular, or internal mammary node recurrence rate was 1%; the chest wall recurrence rate was 5%. These results are significantly less than would be expected in the absence of radiation. Results from other institutions (7–9) also demonstrate that moderate-dose radiation significantly reduces the likelihood of local-regional relapse.

Postoperative radiation clearly reduces local-regional relapse, but its effect on survival remains controversial. In an attempt to address this question, the National Surgical Adjuvant Breast Project (NSABP) conducted a prospective randomized trial in which patients with early breast cancer were treated by radical mastectomy and were then randomly assigned to postoperative radiation to regional lymph node areas, or to a control group (1). The results are shown in Table 3. Postoperative radiation clearly reduced regional relapse, but there was not a statistically significant improvement in survival. Two more recent clinical trials testing the value of postoperative radiation showed similar results (7, 8). The weight of current evidence supports the conclusion that postoperative radiation does not, in general, significantly improve survival. The current understanding is that the large majority of patients with occult residual local-regional

Table 3 Results at 5 years from the NSABP trial testing the value of postoperative radiation therapy

	Relapse-free survival (%)		Regional recurrences (%)	
Histologic axillary nodal status	Irradiated patients	Controls	Irradiated patients	Controls
All negative	78.6	76.0	0	2.1
1–3 positive	49.1	47.8	0	4.5
4 or more positive	28.4	18.0	1.5	12.5

cancer following mastectomy already have occult distant spread. As a result, while postoperative radiation can eliminate the residual local-regional cancer, it does not prevent the appearance of distant metastases or improve survival.

Although postoperative radiation is not of value in unselected patients, further study is required to determine whether or not there are subsets of patients for whom postoperative therapy is of value. There is preliminary evidence that patients with inner or central primaries and positive axillary nodes may be such a subset (7, 10). Many physicians feel that for high-risk patients the prevention of local-regional recurrence is sufficient reason to recommend this treatment. Local-regional recurrence can be highly distressing to a patient, and once manifest is only controlled in 50–67% of cases (11, 12).

With the demonstration that moderate-dose radiation could eradicate subclinical disease in the supraclavicular and internal mammary node regions, it was logical to ask whether or not radiation could eradicate disease in the axilla. MacWhirter, from Scotland, was one of the first to combine total mastectomy with radiation to the axillary, internal mammary, and supraclavicular lymph node areas as a substitute for radical mastectomy. His retrospective results supported the conclusion that more conservative surgery combined with radiation was equivalent to radical mastectomy (13). More recently, the NSABP conducted a clinical trial comparing these two forms of local treatment (14). With an average follow-up time of 36 months, patients with clinically negative axillary nodes treated by radical mastectomy had a local-regional relapse rate of 8%, compared to 5% for patients treated by total mastectomy and regional radiation. Relapse in the axilla was 1% in both groups. The rate of distant relapse and survival was also similar for these two forms of treatment. These results support the use of more conservative surgery combined with radiation as an equivalent to radical surgery.

The next logical question to ask is whether or not radiation can be combined with even more conservative surgery (namely, a resection of only the primary tumor in the breast) and still achieve adequate local-regional tumor control. If moderate doses of radiation can eradicate subclinical disease in the supraclavicular, internal mammary, and axillary lymph node areas, it is reasonable to assume it can eradicate subclinical cancer in the breast. Conservative surgery for resection of gross cancer, combined with moderate-dose radiation to the breast and draining lymph node regions, is what is meant by primary radiation therapy for carcinoma of the breast.

Primary Radiation Therapy

The concept of combining conservative surgery with radiation as a substitute for more radical surgery is not new. Geoffrey Keynes, a surgeon at St.

Bartholomew's Hospital in London, began to treat patients with operable carcinoma of the breast in this fashion as early as 1924 (15, 16). Because of the rudimentary external beam radiation equipment available at that time, Keynes treated patients by radium needle implantation of the breast and draining lymph node areas. A 10-year retrospective review from St. Bartholomew's Hospital was published in 1953 (17). For patients with disease clinically confined to the breast, the 10-year survival rate following simple surgery and radium treatment was 49%, compared to 52% for radical surgery. For patients with clinical spread to axillary nodes, the 10-year survival rate following simple surgery and radium treatment was 27%, compared to 26% for radical surgery. Local-regional control following the more conservative treatment, however, was not as good as that following radical surgery. The local-regional recurrence index was 13% with conservative treatment and 6% with radical surgery. Subsequent experience has shown that implantation of large and complex volumes, such as the breast and draining lymph node areas, cannot provide the radiation dose homogeneity required for optimal local-regional control. Implantation is now reserved for boosts to localized areas, following more homogeneous radiation to those larger volumes using high energy, external beam radiation.

Two early advocates of conservative surgery combined with radiation were Vera Peters and Sakari Mustakallio. Beginning in 1939, Peters, at the Princess Margaret Hospital in Toronto, treated patients with T1 or T2, N0 breast cancers with excision and radiation (18). To compare her results with those achieved by radical surgery, she performed a matched-pair analysis in which each of 184 patients treated by excision and radiation were matched by age, size of primary tumor, and year of treatment to three patients treated by radical mastectomy and radiation. These results carried out to 30 years do not show any significant differences in survival.

Mustakallio, in Helsinki, similarly employed simple excision with radiation in patients with clinically negative axillary nodes (19). He gave radiation therapy considered inadequate by current standards. Using 180–250 kV equipment, he delivered a 2100-rad surface dose in 6 fractions to the breast and lymph node areas: 25% of his patients developed local-regional recurrence by 10 years. Despite this high local recurrence rate, he observed a 5-year survival of 79% and a 10-year rate of 61%. In a more recent review of Mustakallio's experience, Rissanen & Holsti (20) retrospectively compared the results of local excision and radiation, to radical mastectomy and radiation. They found that the 10-year survival rates were similar for the two treatments in patients with T1 primary tumors, but noted an advantage for the more radical surgery in patients with T2 primary tumors.

Based on the findings obtained in retrospective studies such as these, a prospective randomized trial was begun in 1961 at Guy's Hospital in London by Hedley Atkins and John Hayward. Patients aged 50 years or more with T1 or T2, N0 or N1 carcinoma of the breast were randomized to either radical mastectomy and postoperative radiation, or to wide local resection and postoperative radiation. Between 1961 and 1971, 310 patients entered the trial with 188 undergoing radical mastectomy and 182 undergoing wide local resection. This study has been criticized because of the low doses of radiation delivered. Patients undergoing the conservative operation were treated on a 300-kV machine to the axillary and supraclavicular lymph node areas, and received 2500–2700 rad in 12 days. The breast, including the internal mammary chain, was treated on a 6-MeV linear accelerator and received 3500–3800 rad in 3 weeks. While the dose delivered to the breast was close to what is required for control of subclinical disease, the dose to the axilla was extremely low. This is especially true for N1 patients (axillary nodes clinically suspicious), many of whom had gross axillary disease. As a result of these low radiation doses, 22% of patients in the conservatively treated group developed axillary recurrences. In comparison, only 1% of patients treated by radical mastectomy developed axillary recurrences. As noted before, with adequate doses to the axilla, recurrences in that area are uncommon. Of interest is that with the higher dose used to treat the breast, only 5% of patients treated by wide excision and radiation developed recurrence in the residual breast tissue.

The 5-year results from the Guy's Hospital trial were published in 1972 (21) and the updated 10-year results (see Table 4) in 1977 (22). The results are subdivided by whether or not axillary nodes were clinically negative (N0) or positive (N1). N0 patients showed no significant difference in the incidence of distant metastases or in survival, which confirms the results seen at 5 years. N1 patients showed a slight increase in the incidence of distant metastases in the conservatively treated group and a decrease in survival. While the previously published 5-year results were statistically significant, the updated 10-year results were not.

Table 4 The 10-year results from the Guy's Hospital trial

	Axillary nodes clinically negative		Axillary nodes clinically positive	
Therapy	Distant metastases (%)	Survival (%)	Distant metastases (%)	Survival (%)
Wide excision and radiation	44	52	65	30
Radical mastectomy and radiation	46	58	60	43

The results of the Guy's Hospital trial clearly indicate that wide local excision and low-dose radiation do not effectively control local-regional tumors. As will be indicated, the results with more modern radiotherapeutic techniques and adequate radiation dose provide vastly improved local-regional tumor control. Despite the inadequacies in the radiation treatment, the 10-year results actually support the use of conservative treatment, especially in patients with clinically negative axilla, by demonstrating no differences in survival or distant metastases.

In the early 1960s, several institutions in France began treating patients with early breast cancer primarily with radiation. This approach was based on the earlier work of Baclesse, who in the 1930s provided histological evidence of breast cancer eradication following adequate doses of radiation using orthovoltage equipment (23). With the widespread introduction of megavoltage ^{60}Co and ^{137}Cs units in France in the late 1950s, it became possible to treat patients with improved dose homogeneity and with sparing of the uppermost levels of the skin. The experience from these institutions not only provides long-term results following primary radiation, but also illustrates the importance of certain technical factors involved in this treatment.

In 1978, Calle et al (24) at the Curie Foundation in Paris published their results on 514 patients with operable breast cancer. Their treatment protocol depended on the clinical stage of the cancer. If the primary tumor was 3 cm or less, and no suspicious axillary nodes were present, patients were treated by lumpectomy followed by radiation. If the primary tumor was greater than 3 cm, or if suspicious axillary nodes were present, the diagnosis was confirmed by needle biopsy and treatment was given exclusively with external beam radiation. For patients treated by lumpectomy and radiation, the absolute survival, without evidence of disease, was 85% at 5 years and 75% at 10 years. The local-regional recurrence rate in this group was 13% at 5 years; only 2% of patients had poor cosmetic results. In contrast, the results achieved by exclusive radiation were less satisfactory. Of 394 patients, 234 (59%) required secondary surgery either for persistent or recurrent disease. In addition, with the higher doses of radiation required to treat gross disease, the cosmetic result was not as good. Only 38% of patients had good to excellent cosmetic results, and 17% had poor results. Salvage of radiation failures by mastectomy was good, so that survival in this group did not appear to be compromised. The experience from the Curie Foundation provides long-term support for the efficacy of conservative surgery combined with radiation, but indicates that lumpectomy is important for maximizing initial local tumor control. Similar 10-year results were reported by Amalric et al (25) at the Marseilles Cancer Institute.

Pierquin et al (26) at the Henri Mondor Hospital in France recently published their experience with primary radiation therapy in 177 patients with operable breast cancer. Their policy regarding the extent of surgery was similar to that at the Curie Foundation: only those patients with primary tumors less than 3 cm underwent lumpectomy. Of note in this series, however, was the routine use of interstitial implantation. Treatment in all patients began with megavoltage ^{60}Co; 4500 rad in 5 weeks was delivered to the entire breast and draining lymph node areas, and booster doses were delivered to the axillary and internal mammary lymph node areas. An additional 2500–3700 rad was then delivered to the primary tumor site by interstitial implantation. Figure 1 is a schematic illustration of the radiation doses obtained with interstitial implantation, and demonstrates the high dose of radiation that can be localized to a tumor-bearing volume in the breast, with relative sparing of nearby normal tissues, such as lung and skin. Besides the advantage in dose distribution, the constant, low-dose radiation delivered with interstitial implantation may have a greater antitumor effect than the intermittent, high-dose radiation given with external beam treatment. Pierquin et al chose to use ^{192}Ir as the source for interstitial radiation because of its effective energy (340 keV), long half-life (74.5 days), and convenience of administration. They observed a 5% 5-year local-regional recurrence rate in T1 tumors, and an 8% rate in T2 tumors. Overall survival and relapse-free survival at 5 years were comparable to those for similarly staged patients treated by radical surgery at that institution. These results suggest that the use of interstitial implantation is an important technique for optimizing local tumor control.

Beginning in 1973, a prospective randomized clinical trial testing the value of primary radiation therapy was begun at the National Cancer Institute of Italy at Milan. Women with primary tumors less than 3 cm and clinically negative axillary nodes were randomized to either radical mastectomy or to conservative treatment consisting of a wide local resection and full axillary dissection followed by radiation. The radiation consisted of 5000 rad to the residual breast tissue and a 1000-rad boost using orthovoltage to the region of the primary. As of November 1979, 330 patients had been randomized to the conservative treatment and 328 to radical mastectomy. Four patients treated conservatively had developed local-regional recurrence, compared to seven patients treated by radical mastectomy. In this preliminary report, the actuarial probabilities of survival and relapse-free survival (Figure 2) were comparable for the two treatments (31).

Many US institutions now treat women with early breast cancer with primary radiation therapy (27–30). The experience at our own institution is representative of these recent studies, and illustrates the results that can be obtained (32). This experience includes 176 consecutive patients with UICC stage I and II carcinoma of the breast treated between July 1968 and

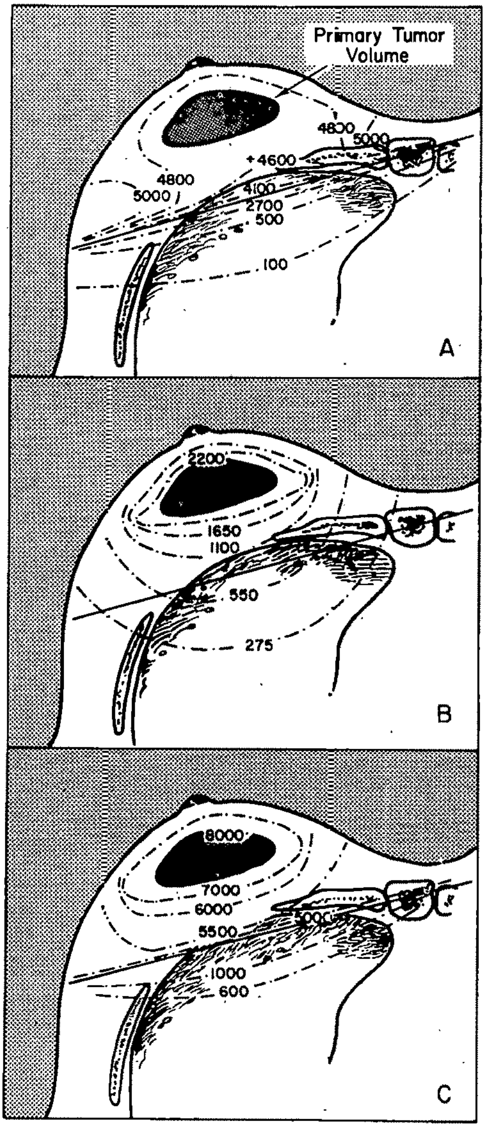


Figure 1 Typical radiation dose distribution seen with external beam plan (*A*), interstitial implant (*B*), and composite (*C*).

December 1977. Eight of these patients had a contralateral breast cancer treated, making a total of 184 breasts treated by primary radiation therapy. In all cases, microscopic examination revealed invasive carcinoma. Patients were intially referred for treatment if they refused mastectomy or were considered medically unfit for surgery. Later in the series, patients were

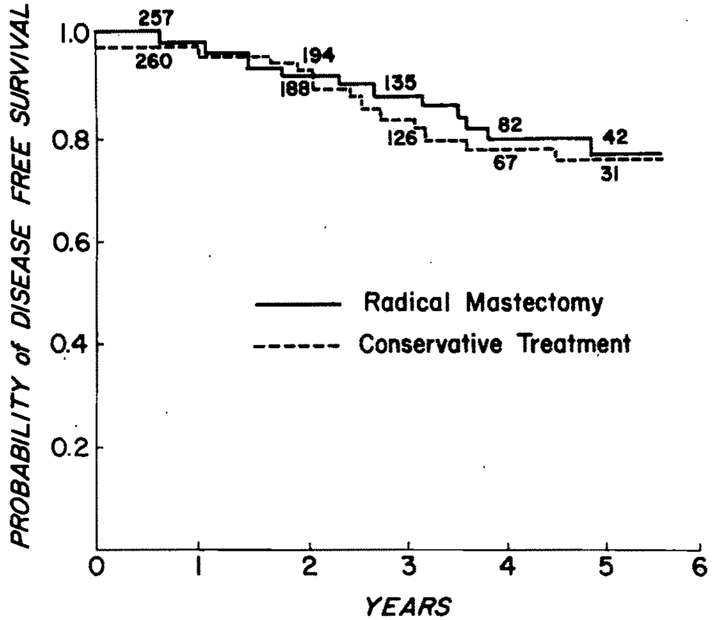


Figure 2 Prospective randomized trial comparing radical mastectomy with conservative surgery and radiation from the National Cancer Institute, Milan, Italy.

referred because of physician or patient preference. Follow-up ranged from 12 to 117 months with a median of 33 months. The details of the therapy have been described previously (33).

In presenting these results, we direct our attention principally to local-regional control rates, because this is the primary objective of local-regional treatment. Since the majority of local-regional failures occur in the first 2–3 years following treatment, these results are considered a good indication of the eventual control rates to be obtained. Of the 184 breast cancers treated, 11 (6%) have had local-regional recurrence. Three of 62 (5%) stage I cancers recurred locally; 8 of 122 (7%) stage II recurred locally. Ten of the failures occurred in the breast alone and one occurred in the breast and axilla.

We have attempted to determine the importance of excisional biopsy and interstitial implantation on local-regional recurrence control (Table 5). Interstitial implantation improved local-regional control ($p<0.05$, by the 2-tail Fisher Exact test). One of 73 patients (1%) treated by interstitial implantation developed a local recurrence, compared to 10 of 111 patients (9%) in whom implantation was not performed. Local-regional control was better in patients having an excisional biopsy than in those having less than an excisional biopsy, but the number of patients failing to have excisional biopsy was small, and the difference in control did not reach statistical

Table 5 Local-regional recurrence rates from the JCRT series

Procedure	Implantation	No implantation	Total
Excisional biopsy	1/63	7/103	8/166 (5%)
Less than excisional biopsy	0/10	3/8	3/18 (17%)
Total	1/73 (1%)	10/111 (9%)	11/184 (6%)

significance (p=0.08). These results corroborate the findings from the Curie Foundation and from Creteil by demonstrating the importance of both excisional biopsy and interstitial implantation in maximizing local tumor control. The actuarial 5-year survival rate in this group of patients was 94% for stage I and 73% for stage II. While 10–15-year survival rates in carcinoma of the breast are still needed, these preliminary results are comparable to those seen following mastectomy in similarly staged patients.

The local control rate in this series is encouraging, but it is not the sole criterion for judging the success of primary radiation therapy. If primary radiation therapy is to be a useful alternative to surgery for early carcinoma of the breast, it must also achieve good cosmetic results. We reported a cosmetic evaluation in which 66% of treated breasts were judged good or excellent by physician evaluation, and 88% were judged good or excellent by the patients themselves (34). Three principal treatment factors influenced the cosmetic outcome: (*a*) the extent and location of the biopsy procedure; (*b*) the time-dose factors of the radiation therapy, and (*c*) the technique of the radiation therapy. Cosmetic results were adversely affected when the biopsy scar was obvious. Increasing doses of external beam radiation above 4500–5000 rad in 5 weeks resulted in greater degrees of retraction and fibrosis of the treated breast. All patients in this series were treated with megavoltage equipment, and permanent skin changes secondary to treatment were infrequent. Interstitial implantation did not diminish the cosmetic outcome.

In our experience, external beam radiation for breast cancer is well tolerated; mild fatigue is the most common side-effect. Treatments are given 5 days a week for approximately 5 weeks. Each treatment takes 15–20 minutes with most of that time spent on patient set-up. Patients rarely compain of nausea, and cranial hair loss is not seen. At the end of 5 weeks, patients may develop local skin changes varying from dry desquamation and erythema, to areas of moist desquamation. These changes are transient and clear within 1–3 weeks. Interstitial implantation is usually performed under general anesthesia and remains in place for approximately 2–3 days. Following recuperation from the anesthesia, patients are up and about their rooms and, typically, do not require analgesics. Late reactions in the irradiated volume can occur in the months following treatment. The most common reaction was rib fractures, which occurred in 10 of the 184 cases (5%). Five patients (3%) developed a nonproductive cough and mild shortness of breath, and chest x rays showed evidence of radiation pneumonitis. In all cases, these symptoms were self-limited and did not require hospitalization. Other transient reactions were seen: one patient developed a pleural effusion following treatment and another patient had paresthesias in her ipsilateral arm. Three patients experienced more significant side-effects. One patient developed marked fibrosis in her treated supraclavicular region and persistent motor weakness in her ipsilateral arm. Another patient developed significant arm edema and a venogram revealed partial obstruction of her innominate vein. The third patient, who presented with a synchronous bilateral breast cancer, developed radiation pericarditis following mediastinal irradiation of her internal mammary nodes. This patient required a pericardiectomy, but is now well.

One of the possible complications of this treatment is the induction of tumors in the irradiated volume. While radiation has been clearly identified as a carcinogen, the risk of carcinogenesis following therapeutic levels of radiation has not been adequately defined. A full discussion of this topic is beyond the scope of this review, except to mention two considerations: the dose of radiation and the age at the time of exposure. The likelihood of tumor induction following therapeutic doses of radiation cannot be extrapolated from estimates following lower doses of radiation. The experience with pelvic irradiation illustrates this point. At one time, pelvic irradiation with doses of 500–1000 rad was used to treat benign uterine bleeding. The follow-up on this experience has shown a three-fold increase in the leukemia risk (35–37). In contrast, higher doses of pelvic radiation (typically combining implantation and external beam radiation) have been used to treat carcinoma of the uterine cervix, and several large studies have failed to demonstrate any increased risk of leukemia in this group of patients (38–40). It has been hypothesized that radiation results in two separate processes: cell transformation, leading to malignancy, and cell inactivation, resulting in loss of clonogenic potential. Current evidence suggests that cell transformation has a linear relation to radiation dose, while cell inactivation has an exponential relation. If this hypothesis is correct, then one would observe an optimal carcinogenic dose of radiation. With higher doses, cell inactivation would be relatively more prominent than cell transformation, and the absolute risk of carcinogenesis would decrease. The available data on the occurrence of leukemia following radiation are consistent with this hypothesis, and indicate that the risk of leukemia following therapeutic levels of radiation is lower than would be predicted on the basis of the experience with lower doses.

A second consideration is age at the time of exposure to radiation. There is evidence from several sources that the relative risk of tumor induction in breast tissue is greater in younger women, than in older women. In a large Massachusetts study examining the risk of breast cancer in women following repeated fluoroscopic examinations of the chest, no increased risk was seen for women 30 years of age or older; while a 3.8-fold increased risk was observed for women aged 15 to 19 (41). Similar findings were noted in another fluoroscopic study from Nova Scotia (42). In a follow-up of women treated at the Radiumhemmet in Sweden for various non-neoplastic conditions of the breast, a four-fold increase in the breast cancer risk was seen. For women aged 34 or less, the risk was 7.1 times greater than expected; for women aged 35 or more, the risk was 2.2 times greater (43). Similarly, the data from the atom bomb experience in Japan indicated an increase in breast cancer in exposed women, the risk being significantly greater for younger women (44).

The available data on tumor induction following radiation for carcinoma of the breast suggest that the risk is small. As noted before, primary radiation therapy for breast cancer began as early as the 1930s, and there have not been any reported second malignancies following this treatment. The experience with postoperative radiation is even more extensive, and with greater follow-up. In a study from Ontario, Canada, the risk of developing a contralateral breast cancer was not increased in patients who received postoperative radiation compared to non-irradiated patients (45). Soft tissue and bone sarcomas have been seen following postoperative radiation. In a 50-year review of records from the Memorial Hospital in New York City, there were nine cases of sarcomas following radiation for breast cancer (46). The total number of irradiated patients is not given, although it was common practice for many years at that institution to deliver postoperative radiation in all patients with positive nodes. For comparison, over a similar time-span at the Columbia-Presbyterian Hospital, a total of 18 cases of lymphangiosarcoma of the ipsilateral arm were seen as a consequence of radical mastectomy (47). Here, too, the total number of patients at risk is not available, so that the likelihood of developing this complication cannot be calculated. For the most part, however, lymphangiosarcoma was considered a rare complication, and was not felt to be a contraindication to radical mastectomy.

The use of adjuvant chemotherapy can follow the same guidelines used following mastectomy. Patients considered for adjuvant chemotherapy in our institution now undergo an axillary sampling prior to treatment. This sampling provides histologic confirmation of axillary node involvement currently required for the use of adjuvant chemotherapy. We recently reviewed our experience with 177 patients undergoing this procedure (48). In most cases a separate incision in the axilla is preferred. Lymph nodes from level one, and sometimes level two, are removed; typically about 10 nodes are recovered. Axillary sampling is a useful procedure with minimal morbidity in those patients for whom adjuvant chemotherapy is contemplated. The timing of these various modalities is currently under review. In most patients with early breast cancer, an excisional biopsy is performed for pathologic confirmation of the diagnosis, and for determination of hormone receptor status. At this point, the various options regarding local treatment are discussed with the patient. If a patient decides on a course of primary radiation therapy, and an axillary sampling is indicated, we have found it convenient and feasible to perform axillary sampling and interstitial implantation of the breast under the same anesthesia. In these patients, the external beam portion of the treatment follows the interstitial implant. Chemotherapy is generally used following completion of the radiation. In some cases chemotherapy has been started prior to institution of radiation. Further study is required to determine the optimal sequencing of the radiation and chemotherapy.

Conclusion

The last two decades have seen considerable evolution in breast cancer treatment, though the optimal treatment of this disease remains uncertain and controversial. From both a conceptual and a practical point of view, breast cancer treatment can be divided into (*a*) local-regional treatment to the breast and draining lymph node areas, and (*b*) systemic treatment for metastatic spread. Mastectomy has been the traditional form of local-regional treatment and has provided local-regional tumor control in 85–90% of cases. Initially, radiation was used following mastectomy. While this experience with postoperative radiation was apparently not of survival value in unselected patients, it clearly demonstrated that moderate doses of radiation can eradicate subclinical deposits of breast cancer. Further study is required to identify whether or not there are subsets of patients in whom postoperative radiation would be of value.

Primary radiation therapy is an alternative to mastectomy that combines conservative surgery for removal of gross disease with moderate doses of radiation for control of subclinical disease. Recent results with this treatment show local-regional tumor control comparable to that achieved by mastectomy. Effective tumor control with radiation depends on the removal of gross disease by surgery and the use of adequate doses and technique of radiation. A booster dose to the involved quadrant of the breast is required for optimal results, and this can be satisfactorily delivered by an interstitial implant. Cosmetic results following primary radiation are typically good to excellent, but, here again, the technique of the surgery and radiation are crucial. Survival results at 5 and 10 years from a number of retrospective studies are also comparable to those of surgery. Side-effects from treatment are minimal when good technique is employed, but more experience is required to judge more fully any late effects of this treatment. The use of adjuvant chemotherapy, for control of metastatic disease, can be combined with primary radiation therapy. Patients considered for adjuvant chemotherapy currently undergo an axillary sampling prior to radiation in order to judge the need for adjuvant chemotherapy. The sequencing of these various modalities of treatment is under review.

Primary radiation therapy is a reasonable alternative to radical surgery for patients who are concerned about mastectomy, and when adequate surgical and radiotherapeutic expertise is available to deliver this treatment.

Literature Cited

1. Fisher, B., Slack, N. H., Cavanaugh, P. J., Gardner, B., Ravdin, R. G. (and Cooperating Investigators). 1970. Postoperative radiotherapy in the treatment of breast cancer: Results of the NSABP clinical trial. *Ann. Surg.* 172:711–30
2. Valagussa, P., Bonadonna, G., Veronesi, U. 1978. Patterns of relapse and survival following radical mastectomy: analysis of 716 consecutive patients. *Cancer* 41:1170–78
3. Baker, R. R., Montague, A. C. W., Childs, J. N. 1979. A comparison of modified radical mastectomy to radical mastectomy in the treatment of operable breast cancer. *Ann. Surg.* 189: 553–57
4. Handley, R. S. 1975. Carcinoma of the breast. *Ann. R. Coll. Surg.* 57:59–66
5. Fletcher, G. H. 1972. Local results of irradiation in the primary management of localized breast cancer. *Cancer* 29:545–52
6. Weichselbaum, R. R., Marck, A., Hellman, S. 1976. The role of post-operative irradiation in carcinoma of the breast. *Cancer* 37:2682–90
7. Höst, H., Brennhovd, I. O. 1981. Does post operative radiation live any place following radical mastectomy? *Cancer.* In press
8. Wallgren, A. 1977. A controlled study: preoperative versus postoperative irradiation. *Int. J. Radiat. Oncol. Biol. Phys.* 2:1167–69
9. Bonadonna, G., Valagussa, P., Rossi, A., Zucali, R., Tancici, G., Bajetta, E., Brambilla, C., De Lena, M., DiFronzo, G., Banfi, A., Rilke, F., Veronesi, U. 1978. Are surgical adjuvant trials altering the course of breast cancer? *Semin. Oncol.* 5:450–64
10. Lacour, J., Bucalossi, P., Cacers, E., Jacobelli, G., Koszarowski, T., Le, M., Rumeau-Rouquette, C., Veronesi, U. 1976. Radical mastectomy versus radical mastectomy plus internal node dissection. *Cancer* 37:206–14
11. Chu, F. C. H., Lin, F. J., Kim, J. H., Huh, S. H., Garmatis, C. J. 1976. Locally recurrent carcinoma of the breast. Results of radiation therapy. *Cancer* 37:2677–81
12. Zimmerman, K. W., Montague, E. D., Fletcher, G. H.: Frequency, anatomic distribution and management of local recurrences after definite therapy for breast cancer. *Cancer* 19:67–74
13. McWhirter, R. 1955. Simple mastectomy and radiotherapy in the treatment of breast cancer. *Br. J. Radiol.* 28: 128–39
14. Fisher, B., Montague, E., Redmond, C., Barton, B., Borland, D., Fisher, E. R., Deutsoh, M., Schwartz, G., Margolese, R., Donegan, W., Volk, H., Konvolinka, C., Gardner, B., Cohn, I., Lesnick, G., Cruz, A., Lawrence, W., Nealon, T., Butcher, H., Lawlon, R. (and other NSABP investigators) 1977. Comparison of radical mastectomy with alternative treatments for primary breast cancer. *Cancer* 39:2827–39
15. Keynes, G. 1929. The treatment of primary carcinoma of the breast with radium. *Acta Radiol.* 10:393–402
16. Keynes, G. 1937. Conservative treatment of cancer of the breast. *Br. Med. J.* 2:643–47

17. Williams, I. G., Murley, R. S., Curwen, M. P. 1953. Carcinoma of the female breast: conservative and radical surgery. *Br. Med. J.* 2:787–96
18. Peters, M. V. 1977. Wedge resection with or without radiation in early breast cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 2:1151–56
19. Mustakallio, S. 1972. Conservative treatment of breast carcinoma—review of 25 years follow up. *Clin. Radiol.* 23:110–16
20. Rissanen, P. M., Holsti, P. 1974. A comparison between conservative and radical surgery combined with radiotherapy in the treatment of breast cancer stage I. A 10-year follow-up study on 866 patients. *Strahlentherapie* 147: 370–74
21. Atkins, H., Hayward, J. L., Kligman, D. J. 1972. Treatment of early breast cancer: A report after 10 years of a clinical trial. *Br. Med. J.* 2:427–29
22. Hayward, J. L. 1977. The Guy's trial of treatments of "early" breast cancer. *World J. Surg.* 1:314–16
23. Baclesse, F. 1949. Roentgen therapy as the sole method of treatment for cancer of the breast. *Am. J. Roentgenol. Radium Ther. Nucl. Med.* 62:311
24. Calle, R., Pilleron, J. P., Schlienger, P., Vilcoq, J. R. 1978. Conservative management of operable breast cancer. *Cancer* 42:2045–53
25. Amalric, R., Santamaria, F., Robert, F., Seigle, J., Altschuler, C., Spitalier, J. M., Brandone, H., Ayme, Y., Pollet, J. F., Abed, R. 1981. Radiation therapy with and without primary limited surgery for operable breast cancer: a 20-year experience at the Marseilles Cancer Institute. In press
26. Pierquin, B., Owen, R., Maylin, C., Otmezguine, Y., Raynal, M., Mueller, W., Hannoun, S. 1980. Radical radiation therapy of breast cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 6:17–24
27. Chu, A. M., Cope, O., Russo, R., Wang, C. C., Schulz, M., Wang, C. A., Rodkey, G. 1980. Treatment of early stage breast cancer by limited surgery and radical irradiation. *Int. J. Radiat. Oncol. Biol. Phys.* 6:25–30
28. Prosnitz, L., Goldenberg, I. S. 1975. Radiation therapy as primary treatment for early stage carcinoma of the breast. *Cancer* 35:1587–96
29. Alpert, S., Ghossein, N. A., Stacey, P., Migliorelli, F. A., Efron, G., Krishnaswamy, V. 1978. Primary management of operable breast cancer by minimal surgery and radiotherapy. *Cancer* 42:2054–58
30. Montague, E. D., Gutierrez, A. E., Barker, J., Tapley, N. D., Fletcher, G. 1979. Conservative surgery and irradiation for the treatment of favorable breast cancer. *Cancer* 43:1058–61
31. Veronesi, U. 1981. Personal communication
32. Hellman, S., Harris, J. F., Levene, M. B. 1980. Radiation therapy of early carcinoma of the breast without mastectomy. *Cancer.* 46:988–94
33. Harris, J. R., Levene, M. B., Hellman, S. 1978. Role of radiation therapy in the primary treatment of carcinoma of the breast. *Semin. Oncol.* 5:403–16
34. Harris, J. R., Levene, M. B., Svensson, G., Hellman, S. 1979. Analysis of cosmetic results following primary radiation therapy for stages I and II carcinoma of the breast. *Int. J. Radiat. Oncol. Biol. Phys.* 5:257–61
35. Alderson, M. R., Jackson, S. M. 1971. Long-term follow up of patients with menorrhagia treated by irradiation. *Br. J. Radiol.* 44:295–99
36. Brinkley, D., Haybittle, J. L. 1969. The late effects of artificial menopause by x-radiation. *Br. J. Radiol.* 42:519–21
37. Doll, R., Smith, P. G. 1968. The long-term effects of x-irradiation in patients treated for metropathia hemorrhagica. *Br. J. Radiol.* 41:362–68
38. Hutchison, G. B. 1968. Leukemia in patients with cancer of the cervix uteri treated with radiation. A report covering the first 5 years of an international study. *J. Natl. Cancer Inst.* JNCI 40:951–82
39. Simon, N., Brucer, M., Hayes, R. 1960. Radiation and leukemia in carcinoma of the cervix. *Radiology* 74:905–11
40. Zippin, C., Bailar, J. C., Hohn, H. I., Lum, D., Eisenberg, H. 1971. Radiation therapy for cervical cancer—late effects on life scan and on leukemia incidence. *Cancer* 28:937–42
41. Boice, J. D. Jr., Monson, R. R. 1977. Breast cancer in women after repeated fluoroscopic examinations of the chest. *J. Natl. Cancer Inst.* 59:823–32
42. Myrden, J. A., Hiltz, J. E. 1969. Breast cancer following multiple fluoroscopies during artificial pneumothorax treatment of pulmonary tuberculosis. *Can. Med. Assoc. J.* 100:1032–34
43. Baral, E., Larsson, L., Mattsson, B. 1977. Breast cancer following irradiation of the breast. *Cancer* 40: 2905–10

44. Tokunaga, M., Norman, J. E., Asano, M., Tokuoka, S., Ezaki, H., Nishimore, I., Tsuji, Y. 1979. Malignant breast tumors amoung bomb survivors, Hiroshima and Nagasaki, 1950–1974. *J. Natl. Cancer Inst.* 62:1347–59
45. McCredie, J. A., Inch, W. R., Alderson, M. 1975. Consecutive primary carcinoma of the breast. *Cancer* 35: 1472–77
46. Kim, J. H., Chu, F. C., Woodward, H. Q., Melamed, M. R., Huvos, A., Cantin, J. 1978. Radiation-induced soft tissue and bone sarcomas. *Radiology* 129:501–8
47. Stout, A. P., Lattes, R. 1967. *Tumors of the Soft Tissue.* Washington, DC: Armed Forces Institute Pathol.
48. Bornick, L. E., Rose, C. M., Harris, J. R., Levene, M. B., Weinstein, M., Kaufman, C., Silen, W., Hellman, S. 1980. Technique and results of axillary sampling in patients undergoing definitive breast irradiation. *Abstr. Proc. Am. Soc. Therapeutic Radiologists.*

Ann. Rev. Med. 1981. 32:405–12
Copyright © 1981 by Annual Reviews Inc. All rights reserved

CONVULSIVE AND DRUG THERAPIES OF DEPRESSION

❖7407

Max Fink, M.D.

Department of Psychiatry and Behavioral Science, School of Medicine, State University of New York at Stony Brook, Stony Brook, New York 11794

Depressive disorders are complex alterations in mood state; they occur at any age and are of unknown and seeming multiple pathogenesis. There is no generally accepted reliable prophylaxis or treatment for depressive disorders. Various therapies are used, ranging from environmental change, reassurance, counselling, and psychotherapy to drug therapies and to such unusual interventions as electroconvulsive therapy (ECT), insulin coma, and leukotomy. Environmental and verbal therapies are generally used for the altered mood states following bereavement, loss of a job or a lover, or other "reactive" depressive states. For more serious mood disturbances, particularly those with endogenous features or suicide threats, specific antidepressant drug therapies are generally recommended. Studies comparing the efficacy of tricyclic antidepressants or monoamine oxidase inhibitors with ECT consistently find ECT to be clinically more effective and, at the least, equivalent in safety (28). Yet, ECT is infrequently prescribed. The reasons are many, including inelegance, expense, risks, side-effects, fear, and the need for special training and special facilities for ECT. But, if the published reports are correct that ECT is an effective and rapidly active treatment, which is also as safe as other therapies, when and how is it to be recommended to overcome the fear and antipathy to its use? Are there guides to the recommendation of ECT that may be useful in clinical medical practice?

Diagnostic Considerations

The best guide in choosing a therapy for depression is the clinical diagnosis combined with estimates of the severity of the illness, the acuteness of onset, and the presence of well-defined symptoms. But the number and variety of

0066-4219/81/0401-0405$01.00

diagnostic classifications of mental illnesses seem infinite, and depressive disorders are in the forefront of the more complex and insecure typologies (46, 57). An enduring and useful distinction is that reflected in the endogenous (psychotic, vital, physiological) and reactive (neurotic, personal, psychological) classification. The endogenous group includes subjects with severely depressed mood, diurnal variation, guilt, retardation, insomnia, anorexia, loss of weight, and loss of libido. The reactive group encompasses those with milder illnesses, prone to fluctuate from day to day and lacking the characteristic vegetative features of the endogenous disorders. This distinction has considerable merit for the decision to use ECT. Other characterizations, such as bipolar-unipolar or anxious-hostile depression, lack sufficient reliability for clinical use. Similarly, the recent American Psychiatric Association approved classification, DSM-III, distinguishes among major depressive state, cyclothymic, dysthymic, and atypical bipolar disorders; but these distinctions have yet to be tested in clinical practice, so that DSM-III remains of limited usefulness in selecting a therapy for a depressed individual.

For convulsive therapy, various predictors have been related to a favorable outcome (Table 1). While none alone is reliable, a patient showing many symptoms with moderate or high severity is more likely to respond well to ECT (and to tricyclic antidepressants) than a patient in whom the symptoms are few and mild. Poor prognostic indicators have also been defined (Table 2). While these indicators are derived from prognostic studies, without direct dependence on a single diagnostic convention, the knowledgeable reader will realize that the favorable criteria define patients with an endogenous depressive syndrome of recent onset with sufficient severity to require hospitalization.

ECT-Antidepressant Drug Comparisons

That depressive psychoses, almost regardless of subtype, are the principal indications for convulsive therapy is a recent development. Induced seizures were introduced in 1934 for the treatment of schizophrenia. By 1938,

Table 1 Indicators of a favorable outcome with ECT

Insomnia	Inability to cry
Early morning awakening	Inability to concentrate
Diurnal mood swing (worse in a.m.)	Delusions of worthlessness, guilt; somatic delusions
Anorexia	Suicidal thoughts, activity
Weight loss	Stupor
Dry mouth	Elevated plasma cortisol
Constipation	Abnormal dexamethasone suppression test
Loss of libido	
Retardation	

Table 2 Indicators of an unfavorable outcome with ECT

Anxiety
Initial insomina
Worsening in p.m.
Fluctuating course
Tearfulness
Hypochondriasis
Hysteria
Paranoia
Mildness of symptoms

reports that ECT was singularly useful in the relief of mania and depression dotted the psychiatric literature. ECT continued to be widely used for many mental conditions until the new psychoactive drugs were introduced in the 1950s, and replaced ECT in most conditions outside severe mood disorders.

The early studies (1938–1950) found severely depressed patients showing greater improvement, fewer relapses, and shorter periods of hospitalization after convulsive therapy than historical controls (including patients' own prior episodes) or patients receiving no therapy. The favorable response of patients with involutional depression was particularly compelling since such patients were usually treated with estrogens, with only limited success.

Between 1945 and 1963, studies of the specificity of the seizures showed greater reductions in symptom scores, better discharge evaluations, and shorter hospitalizations for ECT than for placebo, or simulated ECT (anesthesia alone), or subconvulsive treatments. These experiences set the stage for the ECT and antidepressant drug comparisons that are the basis of our present practice. ECT was found more effective or equal to thymoleptics (19, 27, 35, 36, 39, 50, 52, 55, 58, 62, 68) (Table 3). ECT was also more effective than monoamine oxidase inhibitors (MAOI) (39, 47, 48, 62, 65), or phenelzine and amitriptyline combined (22). These studies have rightfully been criticized for using doses of the MAOI that are now considered too low, and further evaluations are needed. Comparisons of the relative efficacy of ECT and lithium (in depression or mania) or ECT and antipsychotic drugs are lacking.

No significant advantage has been reported for the use of antidepressant drugs combined with ECT over that of ECT alone. This conclusion includes

Table 3 ECT efficacy in psychotic depression

Type of study	Number of studies	Chi^2	p value[a]
ECT vs simulated ECT	5	16.3	5×10^{-5}
ECT vs drug placebo	3	77.9	1×10^{-18}
ECT vs tricyclics	6	25.5	4×10^{-7}
ECT vs MAOI	4	78.8	7×10^{-19}

[a] Favoring ECT.

ECT and imipramine (40, 61, 68), amitriptyline (45), phenelzine (40), chlorpromazine (4), and L-tryptophan (24).

However, subsequent treatment with antidepressant drugs sustains improvement better than sedative drugs (45) or no maintenance drugs (40, 61). The role of lithium is still unclear, but recent reviews are favorable (12, 31, 60).

While the risks of ECT are well defined, those of the antidepressant drugs have only recently achieved notoriety. For ECT, the principal risks are amnesia and an organic mental syndrome, which are frequently observed. But the incidence of these effects may be reduced by such procedures as hyperoxygenation, unilateral nondominant placement of the electrodes, and minimal induction currents. Death, reportedly occurring in fewer than 4 in 100,000 treatments, is a risk comparable to that of anesthesia alone. Other risks, such as fracture, induced fear and panic, spontaneous seizures, and death are hazards that can be prevented by the proper administration of the known treatment procedures.

The risks of the alternate treatments are considerable, but are less well defined. The onset of antidepressant effects is often delayed, requiring 10 –14 days for efficacy. The anticholinergic effects of the tricyclic antidepressants elicit dry mouth and difficulty in urination and in vision, particularly in older patients.. Cardiovascular complications, particularly the incidence of arrhythmia and hypotension, may be severe (10, 13, 16, 37, 43, 49). Hypotension may be so severe as to cause falls and fracture (43). The use of the MAOI is restricted by hypertensive crises, cardiovascular and hepatic toxicity, and their interactions with common foodstuffs, tricyclic drugs, and some drugs in medical practice (37).

In comparisons of the safety of ECT and the antidepressant drugs, Avery & Winokur (7–9) make a strong case for ECT; they found a lower suicide and death rate in ECT-treated patients. This is confirmed by a large Italian study of ECT and imipramine (6) and a study that found ECT effective in reducing deaths by suicide in schizoaffective patients (66).

This evidence of the clinical superiority of ECT over antidepressant drugs is compelling; yet in practice, ECT is reserved for those patients who fail to improve with antidepressant drugs, and for the more severely ill patients where a threat to life is clear. This attitude seems proper, since ECT requires greater skills and special facilities for its administration than do the antidepressant drugs. Therapists must have special knowledge of electrophysiology, anesthesia, electrode placements, and the effects of electric currents. The treatment requires equipment for resuscitation, for the management of cardiopulmonary emergencies, and for EKG monitoring. It should generally be given in a hospital or specialized treatment facility, and the cost of this special care must be calculated. For these reasons—complexity, risk, and expense—ECT is reserved for the severely ill.

Special Considerations

In some clinical situations, however, the use of ECT is favored. Suicide threat, recent suicide attempt, severe inanition and cachexia, hyperpyrexia, and exhaustion compel the early use of ECT. This is particularly true in patients exhibiting exhaustion in mania and catatonia, where drug regimens may require heroic dosages and present difficulties in administration.

Psychotic patients with severe medical illnesses, particularly those with renal, cardiac, or hepatic decompensation, or recent coronary occlusion, may be poor candidates for antidepressant drugs with cardiovascular or autonomic effects. While each seizure in ECT is a demanding experience, eliciting elevations in blood pressure, tachycardia, and arrhythmias, the complication rate is low, and even patients with recent myocardial infarction and those with cardiac pacemakers have been successfully treated (1, 14, 41, 69). ECT has been used successfully in patients with severe parkinsonism (3, 5, 11, 17, 25, 51, 59, 71), in patients with severe diabetes (70), and in a patient with a shunt in place for normal pressure hydrocephalus (67). On balance, for patients with compromising medical conditions, ECT is often the safer and more secure antidepressant therapy.

Pregnancy may complicate the treatment of the psychotic patient. Mania or catatonia during the first half of pregnancy may warrant the use of lithium, antidepressant, or antipsychotic drugs. These compounds cross the placental barrier and the risk of teratogenicity may be a sufficient reason to use ECT. ECT does not harm the fetus (15, 64) nor does it induce miscarriage in patients near term (30, 63).

Some severely depressed patients exhibit delusions of guilt, worthlessness, or somatic concern. Such patients frequently have a poor, inadequate, or delayed response to tricyclic antidepressant drugs, even when plasma levels are monitored (23, 32, 38, 42). Some authors find that such patients respond well to ECT and suggest that the presence of somatic delusions in a patient with endogenous depression may be considered a primary indication for ECT. Other authors, however, question the conclusion and suggest that delusional depressed patients will improve with combined antipsychotic and antidepressant drug therapy (44, 53, 54).

Recent neuroendocrine studies of psychiatric patients have reported abnormal results in tests of thyroid, adrenal, and pituitary functions. Endogenously depressed patients are particularly likely to show elevations in plasma cortisol and the loss of its diurnal rhythmicity; failure of cortisol secretion to be suppressed by exogenous steroids (usually dexamethasone); blunted response of the usual elevation of growth hormone to hypoglycemia, L-dopa, and TSH; and an obtunded response to TSH to a challenge dose of TRH. These tests are now being studied for their diagnostic value. Some recent authors report that cortisol levels and their response to dexamethasone provide a prognostic test for selecting patients and for determining an end-point to a course of antidepressant therapy (2, 18, 20, 21, 26, 34, 56).

These studies, among others, also provide the basis for a neuroendocrine theory of the antidepressant action of convulsive therapy (28, 29). These views and the neurohumoral theories (33) provide the physician with rational explanations of the ECT process, which may be useful in understanding the efficacy in depression, catatonia, anorexia nervosa, and parkinsonism and in explaining its use to patients and their families.

Conclusions

Antidepressant drugs and ECT are both effective therapies of endogenous depression, and some antidepressant and anxiolytic drugs may influence aspects of neurotic depression favorably. While ECT is clearly more effective than tricyclic antidepressants and monoamine oxidase inhibitors, the additional difficulty in administering ECT, the public image of its special hazards, and the ease of administering the drug therapies make their use generally preferred. There are special conditions, however, such as the severity of illness, the risk of death by inanition or suicide, and medical complications (including pregnancy) that lead the prudent physician to recommend ECT.

Literature Cited

1. Abiuso, P., Dunkelman, R., Proper, M. 1978. Electroconvulsive therapy in patients with pacemakers. *J. Am. Med. Assoc.* 240:2499–60
2. Albala, A. A., Greden, J. F. 1980. Serial dexamethasone suppression tests in affective disorders. *Am. J. Psychiatry* 137:383
3. Ananth, J., Samra, D., Kolivakis, T. 1979. Amelioration of drug-induced parkinsonism by ECT. *Am. J. Psychiatry* 136:8
4. Arfwidsson, L., Arn, L., Beskow, J., d'Elia, G., Laurell, B., Ottosson, J.-O., Perris, C., Persson, G., Wistedt, B. 1973. Chlorpromazine and the antidepressant efficacy of electroconvulsive therapy. *Acta Psychiatr. Scand.* 49: 580–87
5. Asnis, G. 1977. Parkinson's disease, depression, and ECT: a review and case study. *Am. J. Psychiatry* 134:191–95
6. Avery, D., Lubrano, A. 1979. Depression treated with imipramine and ECT: the deCarolis study reconsidered. *Am. J. Psychiatry.* 136:559–62
7. Avery, D., Winokur, G. 1976. Mortality in depressed patients treated with electroconvulsive therapy and antidepressants. *Arch. Gen. Psychiatry* 33: 1029–37
8. Avery, D., Winokur, G. 1977. The efficacy of electroconvulsive therapy and antidepressants in depression. *Biol. Psychiatry* 12:507–23
9. Avery, D., Winokur, G. 1978. Suicide, attempted suicide, and relapse rate in depression. *Arch. Gen. Psychiatry* 35: 749–53
10. Ayd, F. 1978. Guidelines for treating cardiac patients with tricyclic and tetracyclic antidepressants. *Intern. Drug Ther. Newsl.* 13:9–12
11. Balldin, J., Eden, S., Granerus, A.-K., Modigh, M. K., Svanborg, A., Walinder, J., Wallin, L. 1980. Electroconvulsive therapy in Parkinson's syndrome with "On-Off" phenomenon. *J. Neural Transm.* 47:11–21
12. Bennie, E. H. 1978. ECT and lithium. (Letter). *Br. Med. J.* 1:578–79
13. Biggs, J. T. 1978. Clinical pharmacology and toxicology of antidepressants. *Hosp. Pract.* 13:79–84
14. Bodley, P. O., Fenwick, P. B. C. 1966. The effects of electroconvulsive therapy on patients with essential hypertension. *Br. J. Psychiatry* 112:1241–49

15. Boyd, D. A., Brown, D. W. 1948. Electroconvulsive therapy in mental disorders associated with childbearing. *J. Missouri Med.* 45:573–79
16. Brewer, C. 1978. Risks of tricyclic antidepressants. *Br. J. Psychiatry* 132: 107–8
17. Brown, G. L. 1975. Parkinsonism, depression and ECT. *Am. J. Psychiatry* 132:1084
18. Brown, W. A., Haier, R. J., Qualls, C. B. 1980. Dexamethasone suppression test identifies subtypes of depression which respond to different antidepressants. *Lancet* 1:928–29
19. Bruce, E. M., Crone, N., Fitzpatrick, G., Frewin, S. J., Gillis, A., Lascelles, C. F., Levene, L., Mersky, H. 1960. *Am. J. Psychiatry* 117:76
20. Carroll, B. J., Curtis, G. C., Mendels, J. 1976. Neuroendocrine regulation in depression. *Arch. Gen. Psychiatry* 33: 1039–44; 1051–58
21. Carroll, B. J., Greden, J. F., Feinberg, M. 1980. Neuroendocrine disturbances and the diagnosis and aetiology of endogenous depression. *Lancet* 1:321–22
22. Davidson, J., McLeod, M., Law-Yone, B., Linnoila, M. 1978. Comparison of electroconvulsive therapy and combined phenelzine-amitriptyline in refractory depression. *Arch. Gen. Psychiatry* 35:639–44
23. Davidson, J., McLeod, M., Kurland, A. A., White, H. L. 1977. Antidepressant drug therapy in psychotic depression. *Br. J. Psychiatry* 131:493–96
24. d'Elia, G., Lehmann, J., Raotma, H. 1977. Evaluation of the combination of tryptophan and ECT in the treatment of depression. I. Clinical analysis. *Acta Psychiatr. Scand.* 56:303–18
25. Dysken, M., Evans, H. M., Chan, C. H., Davis, J. M. 1976. Improvement of depression and parkinsonism during ECT: A case study. *Neuropsychobiology* 2:81–86
26. Dysken, M., Pandey, G. N., Chang, S. S., Hicks, R., Davis, J. M. 1979. Serial postdexamethasone cortisol levels in a patient undergoing ECT. *Am. J. Psychiatry* 136:1328–29
27. Fahy, P., Imlah, N., Harrington, J. 1963. A controlled comparison of electroconvulsive therapy, imipramine and thiopentone sleep in depression. *J. Neuropsychiatry* 4:310–14
28. Fink, M. 1979. *Convulsive Therapy: Theory and Practice.* New York: Raven. 306 pp.
29. Fink, M., Ottosson, J.-O. 1980. A theory of convulsive therapy in endogenous depression: Significance of hypothalamic functions. *Psychiat. Res.* 2: 49–61
30. Forssman, H. 1955. Follow-up study of 16 children whose mothers were given electric convulsive therapy during gestation. *Acta Psychiatr. Neurol. Scand.* 30:437–41
31. Gerbino, L., Oleshansky, M., Gershon, S. 1978. Clinical use and mode of action of lithium. In *Psychopharmacology: A Generation of Progress,* ed. M. A. Lipton, A. DiMascio, K. F. Killam, pp. 1261–75. New York: Raven
32. Glassman, A., Kantor, S. J., Shostak, M. 1975. Depression, delusions and drug response. *Am. J. Psychiatry.* 132: 716–19
33. Grahame-Smith, D. G., Green, A. R., Costain, D. W. 1978. Mechanism of the antidepressant action of electroconvulsive therapy. *Lancet* 1:254–56
34. Greden, J. F., Carroll, B. J. 1979. The dexamethasone suppression test as a diagnostic aid in catatonia. *Am. J. Psychiatry* 136:1199–1200
35. Greenblatt, M., Freeman, H., Meshorer, E., Sharp, M. 1966. Comparative efficacy of antidepressant drugs and placebo in relation to electric shock treatment. In *Biological Treatment of Mental Illness,* ed. M. Rinkel, pp. 574–94. New York: Page
36. Greenblatt, M., Grosser, G. H., Wechsler, H. 1964. Differential response of hospitalized depressed patients to somatic therapy. *Am. J. Psychiatry* 120:935–43
37. Hartshorn, E. A. 1974. Interactions of CNS drugs—antidepressants. *Drug Intell. Clin. Pharmacy* 8:591–606
38. Hordern, A., Holt, H. F., Burt, C. G., Gordon, W. F. 1963. Amitriptryline in depressive cases. *Br. J. Psychiatry* 109: 815–25
39. Hutchinson, J. T., Smedberg, D. 1963. Treatment of depression: A comparative study of ECT and six drugs. *Br. J. Psychiatry* 109:536–38
40. Imlah, N. W., Ryan, E., Harrington, J. A. 1965. The influence of antidepressant drugs on the response of electroconvulsive therapy and on subsequent relapse rates. *Neuropsychopharmacologia* 4:438–42
41. Jauhar, P., Weller, M., Hirsch, S. R. 1979. Electroconvulsive therapy for patient with cardiac pacemaker. *Br. Med. J.* 1:90–91
42. Kantor, S. J., Glassman, A. H. 1977. Delusional depressions: natural history

and response to treatment. *Br. J. Psychiatry* 131:351–60

43. Kantor, S. J., Glassman, A. H., Bigger, J. T., Perel, J. M., Giardina, E. V. 1978. The cardiac effects of therapeutic plasma concentrations of imipramine. *Am. J. Psychiatry* 135:534–38
44. Kaskey, G. B., Nasr, A., Meltzer, H. Y. 1980. Drug treatment in delusional depression. *Psychiatry Res.* 1:267–77
45. Kay, D. W. K., Fahy, T., Garside, R. F. 1970. A seven month double-blind trial of amitriptyline and diazepam in ECT-treated depressed patients. *Br. J. Psychiatry* 117:667–71
46. Kendell, R. E. 1977. The classification of depression: a review of contemporary confusion. In *Handbook of Studies on Depression,* ed. G. D. Burrows, pp. 3–20. Amsterdam: Excerpta Medica
47. Kiloh, L. G., Child J. P., Latner, G. 1960. A controlled trial of iproniazid in the treatment of endogenous depression. *J. Ment. Sci.* 106:1139–44
48. King, P. D. 1955. Phenelzine and ECT in the treatment of depression. *Am. J. Psychiatry* 116:64–68
49. Klein, D. F., Davis, J. M. 1969. *Diagnosis and Drug Treatment of Psychiatric Disorders.* Baltimore: Williams & Wilkins. 480 pp.
50. Kristiansen, E. S. 1961. A comparison of treatment of endogenous depression with electroshock and with imipramine. *Acta Psychiatry. Scand.* 37:179–88
51. Lebensohn, Z. M., Jenkins, R. B. 1975. Improvement of parkinsonism in depressed patients treated with ECT. *Am. J. Psychiatry* 132:283–85
52. McDonald, I. M., Perkins, M., Marjerrison, G., Podilsky, M. 1966. A controlled comparison of amitriptyline and electroconvulsive therapy in the treatment of depression. *Am. J. Psychiatry* 122:1427–31
53. Minter, R. E., Mandel, M. R. 1979. A prospective study of the treatment of psychotic depression. *Am. J. Psychiatry* 136:1470–72
54. Nelson, J. C., Bowers, M. B. 1978. Delusional unipolar depression. *Arch. Gen. Psychiatry* 35:1321–28
55. Norris, A. S., Clancy, J. 1961. Hospitalized depressions: drugs or electrotherapy? *Arch. Gen. Psychiatry* 5:276–79
56. Papakostas, Y., Fink, M., Lee, J., Irwin, P., Johnson, L. 1980. ECT, melancholia and the dexamethasone suppression test. In preparation
57. Paykel, E. S. 1977. Response to treatment and depressive classification. See Ref. 46, pp. 21–48
58. Robin, A. A., Harris, J. A. 1962. A controlled comparison of imipramine and electroplexy. *J. Ment. Sci.* 108:217–19
59. Savitsky, N., Karliner, W. 1953. Electroshock in the presence of organic disease of the nervous system. *J. Hillside Hosp.* 2:3–22
60. Schou, M. 1973. Prophylactic lithium maintenance treatment in recurrent endogenous affective disorder. In *Lithium. Its Role in Psychiatric Research and Treatment* S. Gershon, ed. B. Shopsin, pp. 269–94. New York: Plenum
61. Seager, C. P., Bird, R. L. 1962. Imipramine with electrical treatment in depression—a controlled trial. *J. Ment. Sci.* 108:704–7
62. Shepherd, M. 1965. Clinical trial of the treatment of depressive illness. *Br. Med. J.* 1:881–86
63. Smith, S. 1956. The use of electroplexy (ECT) in psychiatric syndromes complicating pregnancy. *J. Ment. Sci.* 102:796–800
64. Sobel, D. E. 1960. Fetal damage due to ECT, insulin coma, chlorpromazine or reserpine. *Arch. Gen. Psychiatry* 2:606–11
65. Stanley, W. J., Fleming, H. 1962. A clinical comparison of phenelzine and electroconvulsive therapy in the treatment of depressive illness. *J. Ment. Sci.* 108:708–10
66. Tsuang, M. T., Dempsey, G. M., Fleming, J. A. 1979. Can ECT prevent premature death and suicide in "schizoaffective" patients? *J. Affect. Dis.* 1:167–71
67. Tsuang, M. T., Tidball, J. S., Geller, D. 1979. ECT in a depressed patient with shunt in place for normal pressure hydrocephalus. *Am. J. Psychiatry* 136:1205–6
68. Wilson, I. C., Vernon, J. T., Guin, T., Sandifer, M. G. 1963. A controlled study of treatments of depression. *J. Neuropsychiatry* 4:331–37
69. Youmans, C. R., Bourianoff, G., Allensworth, D. C., Martin, W. L., Derrick, J. R. 1969. Electroshock therapy and cardiac pacemakers. *Am. J. Surg.* 118:931–36
70. Yudofsky, S. C., Rosenthal, N. E. 1980. ECT in a depressed patient with adult onset diabetes mellitus. *Am. J. Psychiatry.* 137:100–1
71. Yudofsky, S. C. 1979. Parkinson's disease, depression, and electroconvulsive therapy: A clinical and neurobiologic synthesis. *Compr. Psychiatry.* 20:579–81

Ann. Rev. Med. 1981. 32:413–25
Copyright © 1981 by Annual Reviews Inc. All rights reserved

DIETARY PRECURSORS AND BRAIN NEUROTRANSMITTER FORMATION

❖7408

John D. Fernstrom, Ph.D.[1]

Laboratory of Brain and Metabolism, Department of Nutrition and Food Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

INTRODUCTION

Several years have passed since an article (1) has appeared in the *Annual Review of Medicine* dealing with effects of the diet on brain serotonin formation. Since that time considerably more information has accumulated providing convincing evidence that the ingestion of single meals rapidly influences the formation of *several* brain transmitters, and, consequently, brain functions and disease states thought to be related to the release of these transmitters. In the short review that follows, I summarize the known interactions between dietary precursors and transmitter formation in brain, and the resulting effects of precursors and the diet on relevant brain functions and disease states. The interested reader is encouraged to seek other, more complete reviews of this work (e.g. 2, 3).

SEROTONIN

Serotonin (5HT) is formed in brain neurons from L-tryptophan (see Figure 1). The rate of 5HT synthesis is normally controlled by tryptophan availability (1, 2), not by end-product inhibition (4). Tryptophan hydroxylase, which catalyzes the rate-limiting step, is unsaturated with substrate at normal brain tryptophan levels. Raising or lowering brain tryptophan alters the saturation of the enzyme, and thus the rate of the reaction

---

[1] These studies were supported in part by grants from the National Institute of Mental Health, NIH, and NASA. The author is the recipient of a Research Scientist Development Award (MH00254) from the National Institute of Mental Health.

0066-4219/81/0401-0413$01.00

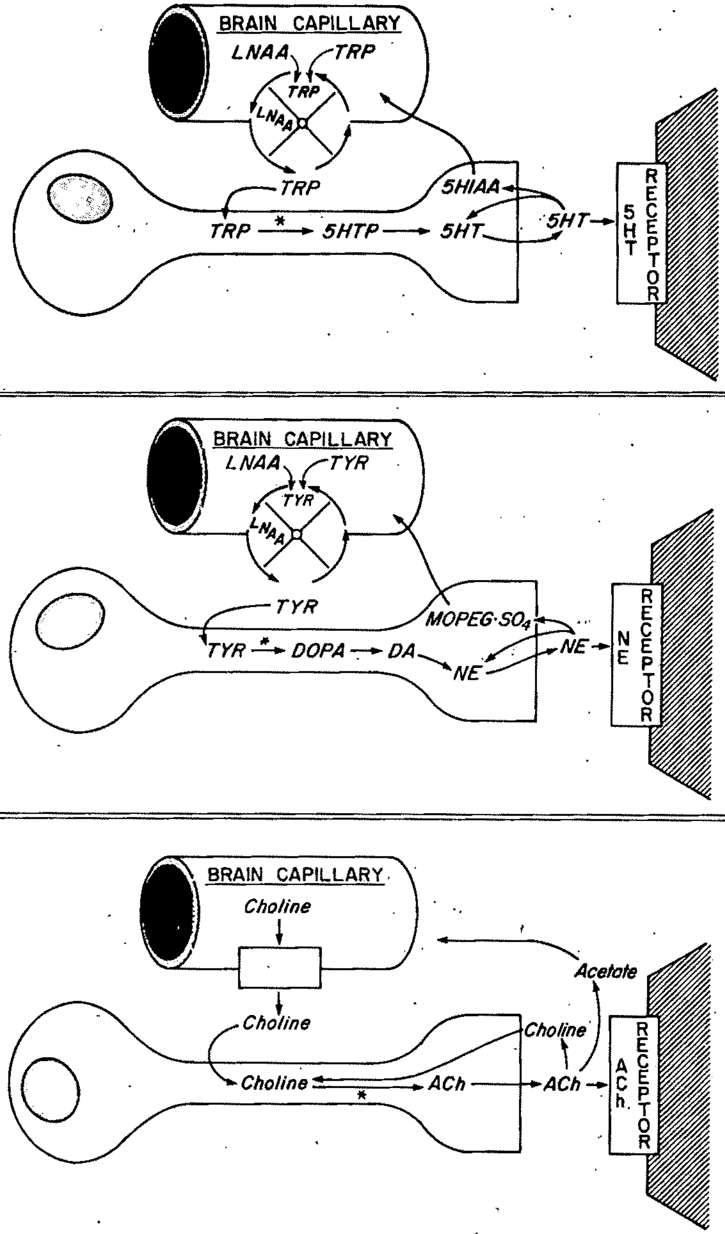


Figure 1 Precursor dependence of neurotransmitter synthesis in the mammalian brain.

(Top) Tryptophan availability and serotonin synthesis. Tryptophan (TRP) is hydroxylated to 5-hydroxytryptophan (5HTP); 5HTP is decarboxylated to serotonin (5HT). Monoamine oxidase initiates the catabolism of 5HT to 5-hydroxyindoleacetic acid (5HIAA). The asterisk

(5). Because of this relationship, any factor that modifies brain tryptophan could potentially influence 5HT formation. One such factor is the rate at which tryptophan is taken up into the brain from the blood. The uptake process depends on the blood level of tryptophan: increasing or decreasing blood tryptophan rapidly raises or lowers brain tryptophan (6). However, tryptophan uptake also depends on the levels in blood of several other large neutral amino acids (LNAA: tyrosine, phenylalanine, leucine, isoleucine, and valine), which are known to compete with tryptophan for access to the uptake carrier (7, 8). Raising their levels in blood reduces brain tryptophan (9); the converse is also true.

Hence, brain tryptophan levels (and 5HT synthesis) can readily be influenced by altering the blood levels of any one or more of the LNAA (which includes tryptophan) (see Figure 1). Over the past several years, we have shown that the ingestion of single meals, which readily alters blood LNAA levels, predictably influences rat brain tryptophan levels and 5HT formation (see 1). Ingestion of a carbohydrate meal, for example, raises brain tryptophan and 5HT; consumption of a protein-containing meal does not modify brain tryptophan or 5HT, even though blood tryptophan increases appreciably. These effects can be explained by the competitive uptake system for LNAA, given the changes induced in the blood levels of the LNAA by the meal. Carbohydrate ingestion causes insulin secretion and lowers blood levels of the aromatic and branched-chain amino acids, while not affecting and often increasing blood tryptophan. Tryptophan in blood thus gains a competitive advantage over the other LNAA, and its uptake into brain is enhanced. Protein ingestion, which causes both insulin secretion and the addition of new amino acid molecules to the blood, increases not only blood tryptophan but also the blood levels of the other LNAA (by an amount proportional to the rise in tryptophan). Tryptophan's competitive standing vis à vis the other LNAA is thus not changed, and no rise in

---

indicates the rate-limiting step in 5HT formation, TRP hydroxylation. TRP competes with other large neutral amino acids (LNAA) for uptake into brain.

(Middle) Tyrosine availability and catecholamine synthesis. Tyrosine (TYR) is hydroxylated to dihydroxyphenylalanine (DOPA); DOPA is decarboxylated to dopamine (DA), which is then betahydroxylated to norepinephrine (NE). Betahydroxylation of DA does not occur in neurons using DA as their transmitter. Monoamine oxidase initiates the catabolism of NE to methoxyhydroxyphenylethyleneglycolsulfate (MOPEG-SO_4), one of the principal NE metabolites in brain. The asterisk indicates the rate-limiting step in catecholamine formation, TYR hydroxylation. TYR competes with other LNAA for uptake into brain.

(Bottom) Choline availability and acetylcholine synthesis. Choline uptake into brain occurs via a mechanism that appears specific for choline. Choline is converted to acetylcholine (ACh) by cholineacetyltransferase (*); this step is rate-limiting, and precursor dependent. See text for further details.

brain tryptophan occurs despite the rise in blood tryptophan (10). For simplicity, we have summarized the changes in blood LNAA that accompany food ingestion as a serum tryptophan ratio (Figure 2). This ratio is tryptophan divided by the sum of the blood levels of the other LNAA [tryptophan/(tyrosine + phenylalanine + leucine + isoleucine + valine)], and accurately predicts brain tryptophan after food consumption: it increases with carbohydrate ingestion, and does not change after protein ingestion (10). Other studies, some utilizing amino acid defined diets, fur-

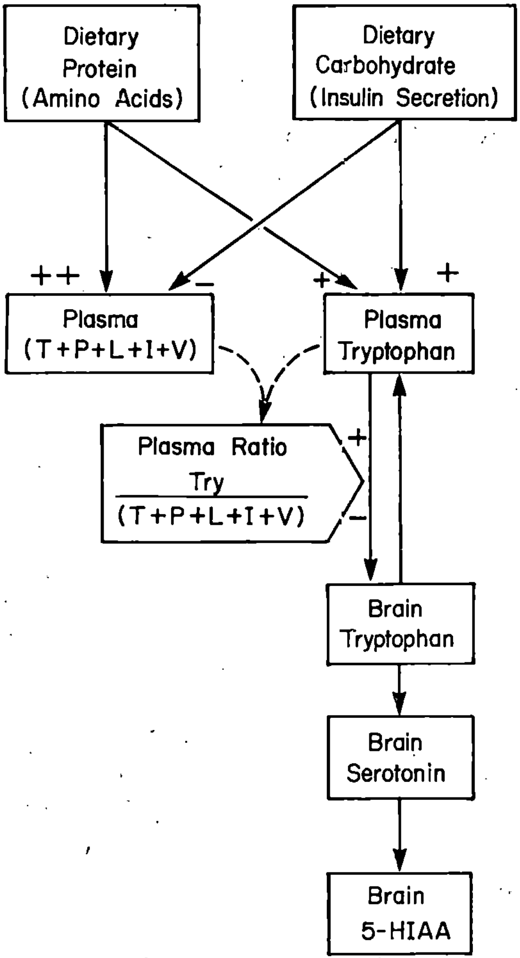


Figure 2 Hypothesized model to describe diet-induced changes in brain serotonin synthesis. The ratio of tryptophan to the sum of other large neutral amino acids [Tyrosine(T) + phenylalanine(P) + leucine(L) + isoleucine(I) + valine(V)] in blood is thought to control tryptophan uptake into brain and brain tryptophan concentration. From (10).

ther support this general hypothesis, i.e. that brain tryptophan (and 5HT) levels follow the serum tryptophan ratio (9).

Since precursor- and diet-induced increases in brain tryptophan elevate brain serotonin synthesis and levels, it might be anticipated that such treatments would also enhance the amount of serotonin released from nerve terminals each time they depolarize. If so, then presumably brain functions and behaviors influenced by serotonin neurons might also be influenced. At present, no data are available on the effects of single meals that alter brain serotonin on such brain functions. However, a growing body of information is extant concerning effects of administered tryptophan. Behaviorally, for example, the administration of tryptophan to animals, either by injection or addition of the pure amino acid to the food supply, improves sleep (11), reduces locomotor activity (12), diminishes the response to acoustic startle (13) and to electroshock (14), suppresses male homosexual behavior (15), attenuates muricide behavior (16), and causes a long-term reduction in dietary protein intake (17). [Other researchers, however, find that tryptophan in combination with fenfluramine diminishes calorie, not protein, intake acutely (18), or that tryptophan alone reduces meal size (19)]. In humans, tryptophan has been observed to act as a hypnotic (11) and as an antidepressant (20–22). Physiologically, tryptophan administration reduces blood pressure in hypertensive rats (23), an effect that can be blocked by prior injection of a serotonin antagonist (e.g. metergoline), or by coadministration of another LNAA. This effect of tryptophan can also be enhanced by coadministration of a serotonin reuptake blocker (fluoxetine). We have also noted in rats that tryptophan injection stimulates significantly natural daily surges in serum growth hormone levels (24) (Figure 3). This latter effect is antagonized by coadministering another LNAA or a serotonin receptor antagonist, and is enhanced by coinjecting a serotonin reuptake blocker. In humans, tryptophan has a modest stimulatory effect on growth hormone secretion (25, 26).

Each of these results is consistent with previous pharmacologic and other findings relating the particular brain function to the activity of serotonin neurons in brain. They suggest indeed that precursor-induced increases in brain serotonin level lead to increased amounts of transmitter impinging on postsynaptic receptors. Perhaps these results with tryptophan will soon lead to similar studies on the effects of short-term, diet-induced changes in brain serotonin on such brain functions and behaviors.

CATECHOLAMINES

Brain neurons synthesize catecholamines (norepinephrine, dopamine, and epinephrine) from tyrosine (see Figure 1). The rates of synthesis of the catecholamine transmitters do appear to be under end-product inhibition

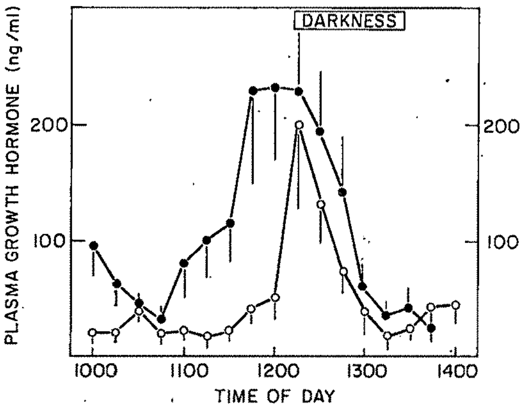


Figure 3 Plasma growth hormone profiles for rats around the onset of the daily dark period (1200 hr). Tryptophan (100 mg/kg) or its vehicle was injected at 0945; each rat received each treatment, in random order, with at least two days intervening between experiments. When animals received vehicle (○—○), typical growth hormone profiles occurred between 1015 and 1345 hr; when animals received tryptophan (●—●) a 97% increase of mean growth hormone levels over vehicle values was observed. Data are expressed as the means ± standard errors; $n = 8$. From (24).

(27), possibly mediated by some sort of neuronal feedback loop, or by direct interaction of transmitter with presynaptic receptors, once released from nerve terminals (28). Even so, it is possible to show that, under appropriate circumstances (e.g. when catecholamine neurons are firing rapidly), dopamine (DA) and norepinephrine (NE) formation and release are also directly influenced by substrate availability, since tyrosine hydroxylase is not fully saturated at normal brain tyrosine levels (29). Thus, tyrosine administration stimulates DA formation in animals pretreated with a DA receptor antagonist (30) or reserpine (31), and NE formation in rats placed in the cold (32), or in otherwise untreated spontaneously hypertensive rats (33). Moreover, food ingestion will modify brain catecholamine synthesis, as a function of its effect on brain tyrosine, in rats placed in the cold [cold increases the brain turnover rate for norepinephrine, thus making the neurons sensitive to tyrosine (32)]. The ability of the meal to have this effect depends on how its ingestion modifies the ratio in blood of tyrosine to the sum of the other LNAA; tyrosine is also an LNAA and thus must compete with the other LNAA for brain uptake [7,8]. In the case of tyrosine, both carbohydrate and protein meals raise the serum tyrosine ratio and brain tyrosine (8, 32).

It is not known why catecholamine neurons must be firing more rapidly to become precursor dependent (i.e. to allow tyrosine administration to stimulate transmitter synthesis). The only available hypothesis argues that tyrosine hydroxylase somehow becomes activated, with the K_m for the cofactor (tetrahydrobiopterin) decreasing greatly (34). As a result, with the cofactor more tightly bound to the enzyme, substrate replaces cofactor as the most limiting component of the reaction mixture, and the precursor effect becomes dominant. Whether or not this particular hypothesis will ultimately be validated is unknown; however, regardless of mechanism, it is clear that in particular animals (hypertensive rats), and under appropriate circumstances (see above), the dependence of NE and DA synthesis on tyrosine availability becomes robust.

What then is the effect on brain function of a precursor-induced increase in catecholamine formation? There is a meager body of data in the literature concerning effects of tyrosine on brain function. This presumably reflects the heretofore generally held belief that only end-product inhibition is significant in controlling catecholamine synthesis. However, quite recently we and others have begun to look for effects of tyrosine on brain outputs that depend on a tyrosine-induced increment in the synthesis and release of either dopamine or NE. To date, the following observations have been made.

1. The injection of tyrosine into spontaneously hypertensive (SH) rats causes a remarkable lowering of blood pressure (Figure 4) (33). This effect is blocked by concurrent administration of another LNAA, valine, which limits tyrosine access to brain; this suggests that the amino acid must gain access to brain to exert its antihypertensive effects. The tyrosine-induced fall in blood pressure is accompanied by an increase in MOPEG-SO_4 levels in the brainstem, but not elsewhere in the central nervous system. MOPEG-SO_4 is reportedly a good index of NE release (35). If so, then the tyrosine-induced increase in its levels may indicate an enhanced synthesis and release rate of NE in brainstem. This correlation is compatible with the general belief that stimulation of brainstem NE receptors by a variety of drugs (e.g. clonidine) lowers blood pressure (see 36). Further work is needed to elucidate fully the mechanism by which tyrosine lowers blood pressure; however, the use of a naturally occurring compound in place of a drug to treat a serious disease is an appealing therapeutic possibility.

2. The oral administration of tyrosine to at least one depressed patient in a double-blind cross-over study had a marked antidepressant effect (37). Perhaps this finding indicates that the subpopulation of patients suffering from organic depression due to inadequte NE release by brain neurons will respond to tyrosine, via a precursor-induced stimulation of NE synthesis and release.

3. The injection of tyrosine into rats treated chronically with reserpine produces an acute reduction in serum prolactin levels, similar to that seen after L-dopa administration (31). The mechanism of this effect may involve tuberoinfundibular neurons that normally synthesize and release dopamine

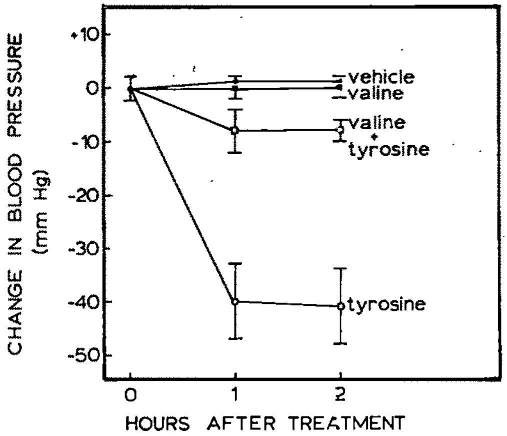


Figure 4 Effect of valine on the antihypertensive action of tyrosine. Tyrosine (100 mg/kg; 0.55 mmol/kg), valine (0.55 mmol/kg), tyrosine plus valine (0.55 mmol/kg of each) or vehicle was administered to groups of four spontaneously hypertensive rats. Blood pressures were measured just before treatment and then at one-hour intervals afterwards. Data are expressed as the change in blood pressure from base line values (means ± standard errors). Tyrosine injection caused a significant fall in blood pressure ($P < 0.01$); treatment with tyrosine plus valine produced a slight decrease in blood pressure ($P < 0.05$), but this effect was not as great as that produced by tyrosine alone. Valine injection had no effect. From (27).

into the pituitary portal circulation. This dopamine is believed to function at the pituitary level as a prolactin release-inhibiting factor (see 38). Reserpine depletes the hypothalamus (and other sites) of dopamine and may indirectly elevate serum prolactin by causing the hypothalamus to release less dopamine, thereby removing this inhibitory influence on pituitary prolactin secretion. Reserpine also stimulates significantly the rate of catecholamine synthesis in brain (39). We suspect that this latter effect renders hypothalamic dopamine neurons responsive to injection of tyrosine; tyrosine has no effect on serum prolactin when injected into otherwise untreated rats (31). If this notion is true, then animals receiving other pharmacologic treatments that lead to enhanced dopamine synthesis and increased serum prolactin [e.g. the administration of a dopamine receptor antagonist (40)] may also respond to tyrosine injection by reducing their serum prolactin levels. Tyrosine may ultimately become a desirable adjunct to drugs such as reserpine or dopamine antagonists, if an association is ultimately proven to exist between high circulating prolactin levels and the incidence of breast carcinoma, as has been suggested (41).

If the use of amino acids such as tyrosine and tryptophan to treat central nervous system disorders ultimately becomes fashionable, the physician must be aware that the consumption of food could drastically diminish the efficacy of the treatment. That such could be the case is most clearly illustrated by studying the effect of diet on the antihypertensive potency of methyldopa (AMD) in SH rats. AMD is a large neutral amino acid analog that is transported into brain by the same competitive carrier system used by the natural LNAA (42). Consequently, it might be anticipated that increased blood levels of the LNAA, caused by any factor, would inhibit AMD uptake into brain. Such is indeed the case (42, 43). The most common natural inducer of an increase in blood LNAA levels is the ingestion of a protein-containing meal. Not surprisingly, therefore, the ingestion of protein prior to AMD injection inhibits the drug's access to brain (43). AMD must enter brain to lower blood pressure (36). Hence, it was predicted and later confirmed that the ingestion of a protein-containing meal would antagonize the antihypertensive action of AMD. This occurs when SH rats consume single meals or when they ingest for long periods diets containing sizeable quantities of protein. As the dietary protein content is increased, the antihypertensive efficacy of AMD is diminished (Figure 5) (43). Dietary protein also reportedly antagonizes another LNAA used therapeutically to treat a disease of the brain; it antagonizes the antiparkinsonian actions of L-dopa (44).

The implications are straightforward and apply regardless of the LNAA used. If a protein meal is consumed prior to amino acid aministration, the

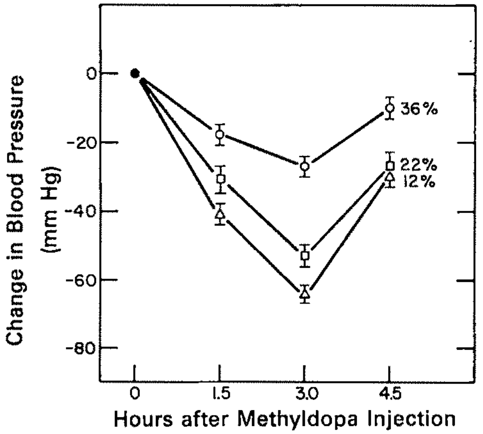


Figure 5 Effect of acute methyldopa administration on blood pressure in spontaneously hypertensive rats consuming protein diets for several weeks. Groups of six hypertensive rats consumed a 12%, 22%, or 36% casein diet for 5 weeks. They then received methyldopa (25 mg/kg), in the morning and blood pressures were measured at the indicated intervals thereafter. By analysis of variance, dietary protein content significantly affected the antihypertensive potency of methyldopa ($P < 0.05$). Mean blood pressures at 0 hr were 208 ± 3 mm Hg. From (43).

efficacy of the amino acid in modifying the brain function under study or treatment will be reduced. In animals or human subjects consuming chronically high levels of protein in their diets, the amino acid will also be less efficacious than in subjects consuming modest (but adequate) amounts of dietary protein. This latter consideration is probably of greatest importance to Americans, who tend to select rather high levels of protein in their diets. To maximize efficacy, then, a physician prescribing a LNAA should thus advise his patient to consume the amino acid between meals, or with carbohydrates (which would cause insulin secretion and lower blood amino acid levels), and should probably also advise a general reduction in dietary protein intake, if the patient consumes substantial amounts daily.

If large neutral amino acids are to be tested in humans for central nervous system effects, it is important to keep the above dietary considerations in mind. Overlooking the relationship of dietary protein content to LNAA efficacy may lead to negative results, where otherwise a positive effect might be obtained.

ACETYLCHOLINE

Acetylcholine (ACh) is synthesized in brain neurons from choline (see Figure 1). The single enzyme involved in this pathway, choline acetyltransferase, appears to be unsaturated with substrate at normal brain choline concentrations. Raising or lowering brain choline accordingly leads to increases or decreases in ACh formation (e g. 45, 46).

Brain choline levels respond readily to changes in blood choline. Choline in blood appears to derive principally from two sources, dietary intake and hepatic synthesis.

The relationship of dietary choline to brain choline levels and ACh synthesis is relatively simple. When the choline content of a meal is sizeable, blood choline levels rise following its ingestion. A large increment in blood choline level rapidly raises brain choline concentrations, which stimulates ACh synthesis (see Figure 1). Similarly, if choline is deleted from the diet, blood and brain choline fall and brain ACh levels decline (46).

The initial work showing that the ingestion of choline influences brain ACh formation involved administering the compound as the chloride salt in the food or by gavage (46). However, choline is normally present in the diet as a constituent of phosphatidylcholine (a lecithin). Hence, it was important to determine if the ingestion of phosphatidylcholine would also raise blood and brain choline levels and stimulate ACh formation. It did (47).

My MIT colleague, Dr. Richard Wurtman, and his associates have pioneered the work on the relationship of choline availability to acetylcholine synthesis (45–47). Together with Dr. John Growdon, as well as other investigators, they have also provided the impetus for investigating the clinical utility of choline and lecithin in the treatment of diseases involving cholinergic dysfunction (see 3). For example, major studies have shown the efficacy of both oral choline chloride and lecithin in treating tardive dyskinesia (see 3 for review). New clinical uses for lecithin are currently under investigation. Perhaps it will be found to be useful in treating memory loss or myasthenia gravis, diseases which respond to drug treatments that increase ACh release.

In contemplating the use of lecithin, one should be aware that most commercially available lecithin preparations contain little choline; they are mostly neutral lipid, not phosphatidylcholine. Hence, commercial preparations of "pure" lecithin provide negligible amounts of choline, and thus do not stimulate ACh formation or improve disease states involving reduced ACh release. When lecithin has been administered to improve such disease states, it has been given as highly purified phosphatidylcholine. Such preparations should soon become more widely available.

SUMMARY

The rates of synthesis of serotonin, acetylcholine, and, under certain circumstances, dopamine and norepinephrine by brain neurons depend considerably on the availability to brain of the respective dietary precursors. This precursor dependence seems to be related to the fact that the enzyme catalyzing the rate-limiting step in the synthetic pathway for each transmitter is unsaturated with substrate at normal brain concentrations. Moreover, brain levels of the individual precursors rise following oral or parenteral administration of the pure compound or the ingestion of certain foods. Precursor-induced increases in brain transmitter formation seem to influence a variety of brain functions and behaviors, which suggests that transmitter release has been enhanced. It now appears that these precursors may become useful as therapeutic agents for the treatment of selected disease states, wherein the disease is related to reduced release of transmitter. Examples are Parkinson's disease (tyrosine), myasthenia gravis (choline or phosphatidylcholine), depression (tyrosine), and possibly abnormal appetite (tryptophan).

Perhaps the future will bring the identification of still other neurotransmitters, whose rates of synthesis depend on precursor availability. Two potential candidates for which some information is already available are glycine (a spinal cord transmitter) and the prostaglandins (some of which may function as neuromodulators or transmitters) (48, 49). Each time a new precursor-product relationship is described, an opportunity becomes available for determining whether the precursor might be useful in treating disease states related to reduced transmitter release by neurons. The opportunities are worth exploring, since the use of a natural dietary constituent, even in purified form, is likely to produce fewer unwanted side-effects than are seen following administration of synthetic drugs.

Literature Cited

1. Fernstrom, J. D. 1974. Modification of brain serotonin by the diet. *Ann. Rev. Med.* 25:1–8
2. Fernstrom, J. D. 1979. The influence of circadian variations in plasma amino acid concentrations on monoamine synthesis in the brain. In *Endocrine Rhythms,* ed. D. T. Krieger, pp. 89–122. New York: Raven
3. Barbeau, A., Growdon, J. H., Wurtman, R. J., eds. 1979. *Nutrition and the Brain,* Vol. 5. New York: Raven. 480 pp.
4. Millard, S. A., Gal, E. M. 1971. The contribution of 5-hydroxyindolepyruvic acid to cerebral 5-hydroxyindole metabolism. *Int. J. Neurosci.* 1:211–18
5. Carlsson, A., Kehr, W., Lindqvist, M., Magnussen, T., Atack, C. V. 1972. Regulation of monamine metabolism in the central nervous system. *Pharmacol. Rev.* 24:371–84
6. Fernstrom, J. D., Wurtman, R. J. 1971. Brain serotonin content: physiological dependence on plasma tryptophan levels. *Science* 173:149–52
7. Pardridge, W. M. 1977. Regulation of amino acid availability to the brain. In *Nutrition and the Brain,* Vol. 1, ed. R. J. Wurtman, J. J. Wurtman, pp. 141–203. New York: Raven
8. Fernstrom, J. D., Faller, D. V. 1978. Neutral amino acids in the brain: changes in response to food ingestion. *J. Neurochem.* 30:1531–38
9. Fernstrom, J. D., Faller, D. V., Shabshelowitz, H. 1975. Acute reduction of brain serotonin and 5-HIAA following food consumption: correlation with the ratio of serum tryptophan to the sum of competing neutral amino acids. *J. Neural. Transm.* 36:113–21
10. Fernstrom, J. D., Wurtman, R. J. 1972. Brain serotonin content: physiological regulation by plasma neutral amino acids. *Science* 178:414–16
11. Hartmann, E. 1977. L-tryptophan: a rational hypnotic with clinical potential. *Am. J. Psychiatry* 134:366–70
12. Marsden, C. A., Curzon, G. 1976. Studies on the behavioral effects of tryptophan and *p*-chlorophenylalanine. *Neuropharmacology* 15:165–71
13. Walters, J. K., Davis, M., Sheard, M. H. 1979. Tryptophan-free diet: effects on the acoustic startle reflex in rats. *Psychopharmacology* 62:103–9
14. Lytle, L. D., Messing, R. B., Fisher, L., Phebus, L. (1975). Effects of long-term corn consumption on brain serotonin and the response to electric shock. *Science* 190:692–94
15. Carruba, M. O., Picotti, G. B., Genovese, E., Mantegazza, P. 1977. Stimulatory effect of a maize diet on sexual behavior of male rats. *Life Sci.* 20:159–64
16. Gibbons, J. L., Barr, G. A., Bridger, W. H. Leibowitz, S. F. 1979. Manipulations of dietary tryptophan: effects on mouse killing and brain serotonin in the rat. *Brain Res.* 169:139–53
17. Woodger, T. L., Sirek, A., Anderson, G. H. 1979. Diabetes, dietary tryptophan, and protein intake regulation in weanling rats. *Am. J. Physiol.* 235:R307–11
18. Wurtman, J. J., Wurtman, R. J. 1977. Fenfluramine and fluoxetine spare protein consumption while suppressing caloric intake by rats. *Science* 198: 1178–80
19. Latham, C. J., Blundell, J. E. 1979. Evidence for the effect of tryptophan on the pattern of food consumption in free feeding and food deprived rats. *Life Sci.* 24:1971–78
20. Coppen, A., Shaw, D. M., Farrell, M. B. 1963. Potentiation of the antidepressive effect of a monoamineoxidase inhibitor by tryptophan. *Lancet* 1:79–81
21. Møller, S. E., Kirk, L., Fremming, K. H. 1976. Plasma amino acids as an index for subgroups in manic-depressive psychosis: correlation to effect of tryptophan. *Psychopharmacology* 49:205–13
22. Walinder, J., Skott, A., Carlsson, A., Nagy, A., Roos, B. E. 1976. Potentiation of the antidepressant action of clomipramine by tryptophan. *Arch. Gen. Psychiatry* 33:1384–89
23. Sved, A. F., Fernstrom, J. D. 1980. Tryptophan administration lowers blood pressure in spontaneously-hypertensive rats. *Fed. Proc.* 39:608

24. Arnold, M. A., Fernstrom, J. D. 1981. L-tryptophan injection enhances pulsatile growth hormone secretion in the rat. *Endocrinology*. In press
25. Koulu, M., Lammintausta, R. 1979. Effect of melatonin on L-tryptophan- and apomorphine-stimulated growth hormone secretion in man. *J. Clin. Endocrinol. Metab.* 49:70–72
26. Glass, A. R., Schaaf, M., Dimond, R. C. 1979. Absent growth hormone response to L-tryptophan in acromegaly. *J. Clin. Endocrinol. Metab.* 48:664–66
27. Spector, S., Gordon, R., Sjoerdsma, A., Udenfriend, S. 1967. End-product inhibition of tyrosine hydroxylase as a possible mechanism for regulation of norephinephrine synthesis. *Molec. Pharmacol.* 3:549–55
28. Cooper, J. R., Bloom, F. E., Roth, R. H. 1978. *The Biochemical Basis of Neuropharmacology*. New York: Oxford. 327 pp. 3rd ed.
29. Wurtman, R. J., Larin, F., Mostafapour, S., Fernstrom, J. D. 1974. Brain catechol synthesis: control by brain tyrosine concentration. *Science* 185:183–84
30. Scally, M. C., Ulus, I. H., Wurtman, R. J. 1977. Brain tyrosine level controls striatal dopamine synthesis in haloperidol-treated rats. *J. Neural. Transm.* 41:1–6
31. Sved, A. F., Fernstrom, J. D., Wurtman, R. J. 1979. Tyrosine administration decreases serum prolactin levels in chronically reserpinized rats. *Life Sci.* 25:1293–1300
32. Gibson, C. J., Wurtman, R. J. 1978. Physiological control of brain norepinephrine synthesis by brain tyrosine concentration. *Life Sci.* 22:1399–1406
33. Sved, A. F., Fernstrom, J. D., Wurtman, R. J. 1979. Tyrosine administration reduces blood pressure and enhances brain norepinephrine release in spontaneously hypertensive rats. *Proc. Natl. Acad. Sci. USA* 76:3511–14
34. Weiner, N., Lee, F. L., Barnes, E., Dreyer, E. 1977. Enzymology of tyrosine hydroxylase and the role of cyclic nucleotides in its regulation. In *Structure and Function of Monoamine Enzymes*, ed. E. Usdin, N. Weiner, M. B. H. Youdim, pp. 109–48. New York: Dekker
35. Korf, J., Aghajanian, G. K., Roth, R. H. 1973. Stimulation and destruction of the locus coeruleus: opposite effects on 3-methoxy-4-hydroxyphenylglycol sulfate levels in the rat cerebral cortex. *Eur. J. Pharmacol.* 21:305–10
36. Van Zwieten, P. A. 1975. Antihypertensive drugs with a central action. *Prog. Pharmacol.* 1:1–63
37. Gelenberg, A. J., Wojcik, J. D., Growdon, J. H., Sved, A. F., Wurtman, R. J. 1980. Tyrosine for the treatment of depression. *Am. J. Psychiatry* 137:622–23
38. Fernstrom, J. D., Wurtman, R. J. 1977. Brain monoamines and endocrine function. In *Reproductive Physiology II*, ed. R. O. Greep, pp. 23–56. Baltimore Univ. Park Press
39. Demarest, K. T., Moore, K. E. 1979. Comparison of dopamine synthesis regulation in the terminals of nigrostriatal, mesolimbic, tuberoinfundibular and tuberohypophyseal neurons. *J. Neural. Transm.* 46:263–77
40. Cocchi, D., Gil-Ad, I., Parenti, M., Stefanini, E., Locatelli, V., Muller, E. E. 1980. Prolactin-releasing effect of a novel anti-dopaminergic drug, domperidone, in the rat. *Neuroendocrinology* 30:65–69
41. Smithline, F., Sherman, L., Kolodny, H. D. 1975. Prolactin and breast carcinoma. *N. Engl. J. Med.* 292:784–92
42. Markovitz, D. C., Fernstrom, J. D. 1977. Diet and uptake of aldomet by the brain: competition with natural, large neutral amino acids. *Science* 197:1014–15
43. Sved, A. F., Goldberg, I. M., Fernstrom, J. D. 1980. Dietary protein intake influences the antihypertensive potency of methyldopa in spontaneously-hypertensive rats. *J. Pharmacol. Exp. Ther.* 214:147–51
44. Mena, I., and Cotzias, G. C. 1975. Protein intake and the treatment of Parkinson's disease with levodopa. *N. Engl. J. Med.* 292:181–84
45. Cohen, E. L., Wurtman, R. J. 1975. Brain acetycholine: increase after sytemic choline administration. *Life Sci.* 16:1095–1102
46. Cohen, E. L., Wurtman, R. J. 1976. Brain acetycholine: control by dietary choline. *Science* 191:561–62
47. Hirsch, M. J., Wurtman, R. J. 1978. Lecithin consumption elevates acetycholine concentrations in rat brain and adrenal gland. *Science* 202:223–25
48. Maher, T. J., Wurtman, R. J. 1979. L-threonine administration increases glycine concentrations in the rat central nervous systems. *Life Sci.* 26:1283–86
49. Vergroesen, A. J. 1977. Physiological effects of dietary linoleic acid. *Nutr. Rev.* 35:1–5

Ann. Rev. Med. 1981. 32:427–34

DOWNWARD TREND IN CARDIOVASCULAR MORTALITY[1]

❖7409

Thomas J. Thom

National Heart, Lung, and Blood Institute, Bethesda, Maryland 20205

William B. Kannel, M.D.

Boston University Medical Center, Boston, Massachusetts 02118

The recent decline in the death rate for coronary heart disease has been the subject of special interest to epidemiologists for the past six years. In the context of its previous epidemic rise and the long-term decline for other cardiovascular diseases, especially stroke, speculation about causes of this dramatic turnaround has stirred interest in the possible impact of medical advances, preventive measures, and lifestyle changes.

MORTALITY TRENDS

The cardiovascular diseases account for almost 980,000 deaths in the United States each year, one of every two deaths. About 640,000 of these deaths are from coronary heart disease and 180,000 from the cerebrovascular diseases (1). Most of these deaths result from the atherosclerotic process, with hypertension as a potent contributing factor. Both atherosclerosis and hypertension often begin early in life, however, so that over 200,000 cardiovascular deaths occur before 65 years of age. Up until the 1960s the number one cause of death even among persons as young as 25–44 years of age was heart disease, primarily coronary heart disease. Heart disease is now only the third leading cause of death in that age group, which suggests that an important beneficial change has taken place since the 1960s in the mortality trend of this disease. That is indeed the case (Figure 1).

---

[1] The US Government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this paper.

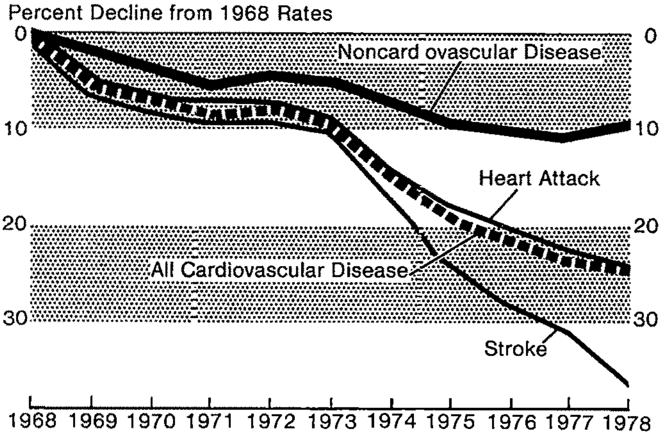


Figure 1 Trends in cardiovascular and noncardiovascular disease: decline by age-adjusted death rates, 1968–1978.

The trend in the number of cardiovascular deaths since 1900 has been upwards primarily because the nation's population of older persons increased, infectious and parasitic diseases came under more control, and the incidence of fatal heart attacks increased at an epidemic rate. When adjusted for the effect of aging of the population, however, rates for the cerebrovascular and hypertensive diseases and for all cardiovascular diseases combined have been declining since at least 1940 (2). The decline in recent years compared to earlier years is unusual in several respects. The death rate for coronary heart disease leveled off in the 1960s and then turned downwards an impressive 21% between 1958 and 1976 (1). Preliminary data from the National Center for Health Statistics (NCHS) show that the decline now exceeds 25% over the past ten years. Secondly, the decline in the death rate for all cardiovascular diseases combined is sharply steeper in recent years, a decline of 2.7% per year so far during the 1970s compared to 0.9% per year in the 1950s and 1960s. Forty-two percent of the decline since 1950 was achieved in just the five years since 1972. For the cerebrovascular diseases, i.e. stroke, the decline exceeds 5% per year since 1972. Thirdly, for the first time all age, race, and sex groups experienced remarkably similar trends for each of the cardiovascular diseases. Declines are not only uniform across these demographic groups, according to the Division of Analysis of the NCHS, they are also fairly uniform for virtually all geographic areas in the nation. There are some slight but important demographic differentials in these recent trends, however. Younger adults and black women are experiencing the steepest coronary declines and rates of decline are greater among older women than among older men. The decline may have begun sooner in certain areas of the country especially among white women. There is a suggestion that life insurance policyholders, probably a higher socioeconomic group than the general population, have even steeper declines (3). For cerebrovascular and hypertensive diseases there are virtually no age or sex differences in steepness of the decline, but declines are steepest for blacks.

Mortality declines are uniformly great for each major cardiovascular disease, with the steepest declines seen for the hypertensive and cerebrovascular diseases. But, with the exception of cancer and the chronic obstructive pulmonary diseases, rates for other natural causes of death are also declining significantly. The importance and uniqueness of declines for the cardiovascular diseases lies in the following facts. These diseases account for half of all deaths, so their decline contributes appreciably to the number of deaths that are delayed until later in life and to the consequent improvement in life expectancy observed in recent years. Secondly, these declines coincide with beneficial changes in cardiovascular risk factor levels, increased efforts in primary and secondary prevention, and expansion of treatment efforts. Thirdly, the decline for coronary disease reverses the earlier epidemic rise. Fourthly, the coronary decline in the United States exceeds declines seen in other countries, many of which are still experiencing an increasing rate (4,5).

Given these declining trends, especially the turnaround in the coronary death rate, have there been concomitant improvements in prevalence, incidence, case-fatality, severity, or survivorship? Is the coronary mortality decline greater for sudden, out-of-hospital deaths? Unfortunately, clear answers to these questions have not emerged. Unpublished data from the Division of Analysis show no definitive trend in prevalence for any cardiovascular disease between health interview surveys in 1972 and 1978. In a Mayo Clinic study the incidence of strokes had been declining since the 1940s at a rate that averaged 3.1 per 100,000 people per year and now (1971–1974) it is declining about 5.3 per 100,000 per year (6). In the Framingham Heart Study the incidence of strokes declined among middle-aged women but not among men (7). The trend in the incidence of coronary attacks is downwards in the Kaiser-Permanente medical care system but no change is observed among Mayo Clinic patients or among subjects in Framingham (7–9). In-hospital case-fatality rates for acute myocardial infarction, available from NCHS, are declining as are lengths of stay. But there is no assurance that these changes are real; they may be due to changes in criteria for admission. Inadequate information exists about any recent changes in survivorship and severity or about trends in sudden compared to nonsudden coronary deaths.

Thus there are no adequate statistics to ascertain whether the cardiovascular mortality declines reflect a reduction in the frequency of cases (initial or subsequent) or that life expectancy after an attack has improved. Yet, one or both improvements have probably taken place because the consensus of opinion is that the recent steep declines are real and not a result of artifacts in the statistics (10–14). Evidence that incidence or severity declined would suggest that preventive measures or lifestyle changes have contributed to the mortality decline. Evidence that life expectancy following attacks increased would suggest that the "explanations" lie in the area of improved medical care. In the absence of such evidence one can only speculate about causes by comparing the timing, extent, and demographic patterns of these mortality trends with the timing, extent, and patterns of trends in risk factors, lifestyle habits, and medical care.

EXPLANATIONS OF THE DECLINE

Probably the first indication in the literature that the national death rate for coronary heart disease was declining rather than rising appeared in 1974 (15). Ever since that article investigators have noted that the decline coincided with beneficial changes in lifestyle and coronary risk factors (7, 11–18). But, inconsistencies remain. Between 1965 and 1978 the proportion of cigarette smoking decreased 35% among men and 10% among women (19). But the proportion of heavy smokers did not decline and declines for women are small whereas their coronary death rate declined faster than for men. Also, coronary mortality declines occurred at older ages where coronary risk appears unrelated to smoking.

During the 1960s and the early 1970s, prevalence of high serum cholesterol ($\geq$ 360 mg) declined 12.5% among men and 22.5% among women (19). But, declines at most levels of cholesterol were modest and were generally not statistically significant. Between 1963 and 1977 per capita consumption of milk, butter, eggs, and animal fats and oils declined 23%, 36%, 15%, and 49% respectively (20, 21). But, meat consumption and total fat content increased. Between approximately 1960 and 1974, the proportion of overweight persons (20% over desirable weight) declined 3.4% for men and 5.2% for women (22). But, actual average weights of men increased during that time. While it cannot be demonstrated that the prevalence of hypertension declined, in recent years the proportion of hypertensive persons who are aware of their condition increased by perhaps 40% and the proportion on treatment almost doubled (23). These changes can be attributed in large part to the National High Blood Pressure Education Program of the National Heart, Lung, and Blood Institute (17). But, it is not clear why prevalence has not declined in the face of long-term use of antihypertensive medication. Moreover, use of those medications began during the rise in the coronary death rate, and the decline in the stroke death rate began before their use. Finally, and without being able to quantify the trend, the jogging and physical activity fad has undoubtedly increased since the 1960s. But most of the big increase was relatively recent, after the presumed start of the coronary death rate decline.

Most of these national trends are also observed in the Framingham Heart Study, which since 1948 has had a representative cohort of the general population of Framingham, Massachusetts, under close surveillance for development of cardiovascular diseases. During that entire time the study has had uniform case ascertainment, risk factor measurement, and criteria for disease end points, all under the same senior investigators. This provides the basis for an examination of some of the time trends in risk factors and cardiovascular morbidity and mortality (7). Consistent with national trends for smoking there has been a marked fall in the proportion of men who smoked. Smoking has actually increased among women. Serum cholesterol levels declined among women but less so among men. Women's weights have fallen sharply and men have gained weight. The prevalence of hypertension declined for women but not for men, which is consistent with the increased rise of antihypertensive treatment among women more than among men. The prevalence of ECG evidence of left ventricular hypertrophy, a powerful and ominous correlate of hypertension, has been decreasing. Most of these changes in Framingham are consistent with a downward trend in cardiovascular disease mortality observed in the study and with a decline in incidence of strokes among women. But, the incidence of strokes in Framingham has not declined among men and the incidence of heart attacks among both men and women has apparently not changed.

While trends observed in coronary risk factors are not perfectly consistent with trends observed in coronary mortality, many investigators are of the opinion that as a means of primary prevention, risk factor modification has taken place and has likely contributed to the recent change in the coronary death rate (11–18, 24). Much of this opinion, which was expressed at the Conference on the Decline in Coronary Heart Disease Mortality held in November, 1978, and sponsored by the National Heart, Lung, and Blood Institute, follows from evidence about risk factors observed in the Institute's Framingham Study. Each 10 mm Hg decrement in blood pressure is accompanied by a 30% corresponding change in the incidence of cardiovascular disease (7). A 10% change in serum cholesterol is associated with a similar 30% difference in coronary incidence. Those who give up the cigarette habit have been found to have only half the coronary mortality of those who continue to smoke. A 10% increase in weight may result in a 30% increase in coronary incidence in men, largely as a result of a worsening in blood lipids and blood pressure. For women, a 10% reduction in weight could result in a 20% reduction in incidence of this disease. It is thus possible that even modest changes in multiple risk factors could be responsible for a substantial decline in coronary death rates. Of these factors the evidence that hypertension control has contributed to the decline in mortality is especially strong for several reasons: (*a*) because high blood pressure is one of the major risk factors for both coronary disease and stroke; (*b*) because hypertensive-related death rates have been on the longest and now the steepest decline; (*c*) because declines are seen in stroke incidence especially for women, the demographic group showing the most improvement in hypertension awareness and treatment; (*d*) because such awareness and treatment increased dramatically during the recent acceleration of the mortality declines; and (*e*) because a recent community-based trial of the effectiveness of an antihypertensive treatment program demonstrated superior mortality reductions among treated hypertensives compared to controls (25). Even though hypertension control appears to be the strongest explanation for the decline in cardiovascular disease mortality, our ability to attribute some or all of the decline to risk factor modification may be more "guilt by association" than a causative explanation.

It is also plausible to associate the coronary decline with the proliferation of coronary care units (CCU), cardiac surgery, and cardiopulmonary resuscitation (CPR) training, with the introduction of Medicare and Medicaid, and with other improvements in medical care. The first CCU was in operation in 1963 and now most coronary patients are treated in them. Although evidence of their benefits is not perfectly conclusive, it does suggest that these units have reduced in-hospital case-fatality rates by about one third (13). The overall impact of these units on the coronary decline must be small, however, since fully two thirds of all coronary deaths occur prior to care in a hospital (13). The major recent innovation in cardiac surgery, the coronary by-pass operation, also is thought to have contributed to the decline (24). However, its impact must be small and does not explain the decline observed. Recent expansion in emergency medical care and CPR have undoubtedly saved lives in selected areas of the country, but these, too, probably had a small impact on the nationwide coronary death rate decline. Improvements in general medical care and access to it have certainly been made, especially since the inception of Medicare and Medicaid. Taken together, all of these changes are thought to have contributed to the turnaround in the coronary mortality trend and the increased steepness of the declining stroke trend. However, the extent of that contribution cannot be properly estimated from available information (13, 14, 24).

IMPLICATIONS OF THE DECLINE

Although causes of the cardiovascular mortality decline remain speculative, these declines are real, substantial, and sustained; they are greater in this country than elsewhere; they accelerated during the past decade of increased efforts towards "healthier" lifestyles; they are observed among all demographic groups; and they are accompanied by declines in noncardiovascular diseases. These patterns suggest that risk factor and lifestyle influences have had a greater impact than specific effects of medical intervention. Although estimates of the relative contributions of these changes require information on the incidence of coronary and stroke attacks and survivorship estimates, the risk factor and lifestyle influences taken as a whole could account for most of the cardiovascular decline. From the observed coincidental changes in the level of risk factors associated with the decline, it is possible to estimate the size of the deline that could result using the Framingham multiple logistic (13). In middle-aged men, at least, the combined impact that could result from the joint reduction in cardiovascular risk factors observed could have produced the amount of decline that has occurred. This assumes, however, that reducing risk factors bestows the same benefit as their having always been at the lower level.

Whatever the explanation for the decline in cardiovascular mortality, one hopes that it will continue. That it will continue is a distinct possibility because many industrialized countries now have considerably lower rates. If the current decline is indeed due to changes in lifestyle, evidently these have not been too drastic to be unacceptable to the American public, and small changes appear to pay large dividends.

Literature Cited

1. Department of HEW, National Center for Health Statistics. 1979. *Chartbook for the Conference on the Decline in Coronary Heart Disease Mortality.* 42 pp.
2. Moriyama, I. M., Krueger, D. E., Stamler, J. 1971. Major cardiovascular-renal diseases. In *Cardiovascular Diseases in the United States,* pp. 268–315. Cambridge: Harvard Univ. Press. 496 pp.
3. Metropolitan Life Insurance Company. 1979. Trends in mortality by age and sex: Metropolitan Life Insurance Company policyholders and US white population, 1969–1977. *Statistical Bulletin,* April-June, 16 pp.
4. Epstein, F. E., Pisa, Z. 1979. International comparisons in ischemic heart disease mortality. In *Proc. Conf. on Decline in Coronary Heart Disease Mortality,* ed. M. Feinleib, R. J. Havlik, pp. 58–90. Washington, DC: Dept. Health, Educ., Welfare, Natl. Inst. Health. 399 pp.
5. Byington, R., Dyer, A. R., Garside, D., Liu, K., Moss, D., Stamler, J., Tsong, Y. 1979. Recent trends of major coronary risk factors and CHD mortality in the United States and other industrialized countries. See Ref. 4, pp. 340–79
6. Garraway, W. M., Whisnant, J. P., Furlan, A. J., Phillips, L. H., Kurland, L. T., O'Fallon, W. M. 1979. The Declining incidence of stroke. *N. Engl. J. Med.* 300(9):449–52
7. Kannel, W. B., Thom, T. J. 1979. Implications of the recent decline in cardiovascular mortality. *Cardiovasc. Med.* 4(9):983–97
8. Friedman, G. D. 1979. Decline in hospitalization for coronary heart disease and stroke: the Kaiser-Permanente experience in northern California, 1971–1977. See Ref. 4, pp. 109–15
9. Elveback, L. R. 1979. Coronary heart disease in Rochester, Minn. 1950–1975: incidence and survivorship. See Ref. 4, pp. 116–22
10. Rosenberg, H. M., Klebba, A. J. 1979. Trends in cardiovascular mortality with focus on ischemic heart disease: United States, 1950–1976. See Ref. 4, pp. 11–14

11. Cooper, R., Stamler, J., Dyer, A., Garside, D. 1978. The decline in mortality from coronary heart disease, USA, 1968–1975. *J. Chron. Dis.* 31:709–20
12. Gordon, T., Thom, T. 1975. The recent decrease in CHD mortality. *Prevent. Med.* 4:115–25
13. Stern, M. P. 1979. The recent decline in ischemic heart disease mortality. *Ann. Intern. Med.* 91:630–40
14. Feinleib, M., Thom, T., Havlik, R. J. 1980. Decline of coronary heart disease mortality in the United States. *Artherosclerosis Reviews,* Vol. 5. New York: Raven In press
15. Walker, W. J. 1974. Coronary mortality: What is going on? *J. Am. Med. Assoc.* 227(9):1045–46
16. Soltero, I., Liu, K., Cooper, R., Stamler, J., Garside, D. 1978. Trends in mortality from cerebrovascular diseases in the United States, 1960 to 1975. *Stroke* 9(6):549–58
17. Levy, R. I. 1978. Progress in prevention of cardiovascular disease. *Prevent. Med.* 7:464–75
18. Blackburn, H. 1980. Why the American decline in coronary heart disease? *Lancet* 1:8161:183–84
19. Department of HEW, National Center for Health Statistics. 1980. *Health, United States, 1979.* DHEW Publ. No. (PHS)80–1232. Washington, DC: GPO. 231 pp.
20. Department of Agriculture. 1978. *Agriculture Statistics 1978.* Washington, DC: GPO. 605 pp.
21. Department of Agriculture. 1976. Food consumption, prices, expenditures. Suppl. to *Agriculture Economic Report.* No. 138. Washington, DC: GPO. 76 pp.
22. Department of HEW, National Center for Health Statistics. 1979. Overweight adults in the United States. *Advance Data.* No. 51. Hyattsville, Md. 11 pp.
23. Department of HEW. 1979. *Healthy People: The Surgeon General's Report on Health Promotion and Disease Prevention; Background Papers.* Washington, DC: GPO. 484 pp.
24. Feinleib, M., Havlik, R. J. 1979. Summary of the conference on the decline in coronary heart disease mortality. See Ref. 4, pp. XXIII–XXVII
25. Hypertension Detection and Follow-up Program Cooperative Group. 1979. Five-year findings of the hypertension detection and follow-up program. *J. Am. Med. Assoc.* 242(23):2562–77

Ann. Rev. Med. 1981. 32:435–42
Copyright © 1981 by Annual Reviews Inc. All rights reserved

PHYSIOLOGY OF CARDIOPULMONARY RESUSCITATION[1]

❖7410

Myron L. Weisfeldt, M. D. and Nisha Chandra, M. D.,

Peter Belfer Laboratory for Myocardial Research and the Cardiology Division,
The Johns Hopkins Medical Institutions, Baltimore, Maryland 21205

INTRODUCTION

For 20 years external sternal compression has been used to create artificial circulation in patients following cardiac arrest (15), and the mechanism for the resulting movement of blood to the brain was thought to be understood. The arrested heart was viewed as a rubber ball filled with fluid with one-way valves at its entrance and exit. The heart is situated between the sternum and the vertebral column, thus external chest compression moves the sternum and was thought to compress the heart against the vertebral column. Such compression, like internal cardiac massage, moved blood from the left ventricle into the aorta as the aortic valve opened. Retrograde flow was prevented by mitral valve closure. Following the period of compression the left ventricle filled with blood when the mitral valve opened.

This widely held concept appeared inconsistent with a number of observations in animal models (21) and man (17), but no alternative mechanisms for the movement of blood were suggested. In the dog arterial and venous pressures during chest compression were similar if not identical (21). In fact, the high intrathoracic venous pressure was considered by some (17, 21) to point to an important injurious effect of "external massage." It was assumed that the very high intrathoracic venous pressures were transmitted peripherally to the brain and caused brain injury.

Only recently did it become clear that other mechanisms beside cardiac compression operate during chest compression. These new suggestions (20)

---

[1] Supported by Grant No. P50-HL-17655-06 from the National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland.

0066-4219/81/0401-0435$01.00

are in part based on confirmation and extension of the earlier observations of the similarity of all intrathoracic vascular pressures during sternal compression.

An important stimulus to us in our studies was the work of Dr. J. Michael Criley and associates (8) dealing with "cough CPR." They demonstrated that by continuous coughing, an arrested patient could maintain himself in the conscious state as long as cough was continued. One ingredient in the cough is clearly a rise in intrathoracic pressure. If this is the active factor in moving blood during cough, then it is a potent mechanism for the movement of blood to the brain.

ANIMAL STUDIES

In extensive hemodynamic studies (20) in large (20–40 kg) dogs we found that during chest compression following cardiac arrest there is essentially an equal rise in central venous, right atrial, pulmonary artery, and aortic pressures. Diastolic pressures were also similar but slightly higher in the aorta than in the right atrium. These pressures in intrathoracic vascular structures were nearly equal to general intrathoracic pressure as indexed by esophageal pressure. It was also clear that forward carotid flow occurred only during chest compression.

We were seeking to identify the pump within the chest responsible for moving blood through the circulatory system against the peripheral vascular resistance, but we could not find one. In any fluid-filled system with an appreciable resistance to flow (such as that present in the peripheral vascular bed), there must be a pressure gradient across the resistance to allow the fluid to flow through it. If there is a pump responsible for the generation of this pressure gradient across the resistance, there must be a pressure gradient across the pump itself. Of course, normally when the heart is acting as a pump there is a substantial pressure gradient between the aorta and the central veins. The lack of such a pressure gradient across the heart during chest compression eliminates the heart or compression of the heart as the sought-for pump.

Hence we tried to identify the mechanism for the generation of the peripheral arterial-venous pressure gradient that *must* be present across the resistance bed if there is to be forward flow of blood during cardiopulmonary resuscitation. By measuring peripheral pressures in the extrathoracic carotid artery and the extrathoracic jugular vein, we confirmed that there was a pressure gradient across this resistance bed. As seen in Figure 1 intrathoracic arterial pressure in the aorta was transmitted relatively completely into the extrathoracic arterial bed; the intrathoracic venous pressure was not transmitted into the extrathoracic bed. This differential transmission of

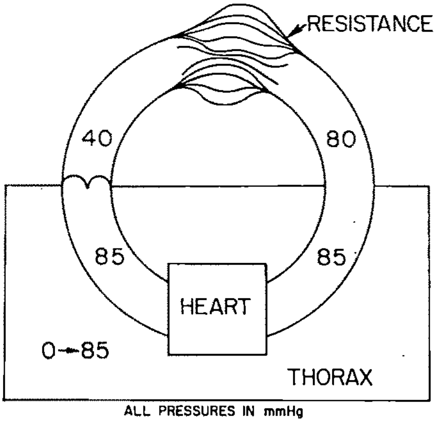


Figure 1 Representative pressures recorded during conventional CPR with forward carotid flow. Pressures are those recorded during compression. Intrathoracic pressures were indexed from esophageal pressures. There is no significant pressure gradient across the heart. The extrathoracic arterial pressure is similar to the intrathoracic aortic pressure. The extrathoracic venous pressure is markedly lower than the intrathoracic venous (right atrial) pressure. There is an extrathoracic arterial-venous pressure gradient that results in forward flow.

the intrathoracic vascular pressure generated the extrathoracic arterial-venous pressure gradient required for forward flow.

Peripheral Arterial-Venous Pressure Gradient

At least three general mechanisms are now thought to contribute to the generation of the extrathoracic arterial-venous pressure gradient observed during conventional CPR in the dog. These factors are (*a*) various venous valving mechanisms are operating, (*b*) peripheral venous capacitance is greater than arterial capacitance, and (*c*) arterial resistance to collapse is greater than venous resistance.

The most easily understood of these mechanisms is that of a valving mechanism. Veins at the thoracic inlet and other veins leading from the brain appear to have anatomic valves that prevent retrograde flow of blood during increases in intrathoracic pressures (9, 10, 16, 18). The valves along the extrathoracic veins, and the one at the thoracic inlet appear to be important.

The second factor contributing to the generation of the peripheral arterial-venous pressure gradient is that venous capacitance is greater than arterial. Clearly, if the same amount of blood were to move from the intrathoracic arterial and from the intrathoracic venous systems into the extrathoracic arterial and venous systems, arterial pressure would rise more than venous pressure because of the differences in extrathoracic arterial and venous capacitance.

The third contributor to the peripheral arterial-venous pressure gradient is the difference in arterial and venous resistance to collapse. Venous structures readily collapse when inside pressures fall below surrounding pressures by even a small amount. Recently we showed that the in vivo carotid artery at the thoracic inlet exhibits considerable intrinsic resistance to collapse. This resistance to collapse can be increased in the presence of vasoconstrictor agents (23). Resistance to collapse of the arterial vessels would allow blood flow to continue toward the brain despite surrounding intrathoracic pressures which exceed intravascular pressures. The veins would readily collapse at the exit to the high pressure region, i.e. at the thoracic inlet (1, 13).

Blood returns from the periphery to the central circulation between compression cycles. Extrathoracic venous pressure rises (20) when blood flows from arteries to veins during compression. Between compressions intrathoracic pressure falls to near atmospheric, and an extrathoracic-to-intrathoracic venous pressure gradient appears, which leads to flow into the chest. Right heart and pulmonary blood flow is also diastolic, at least in part (7). With conventional CPR, right heart compression may be a component of the mechanism for pulmonary flow.

Movement of Blood Through Direct Cardiac Compression

As indicated previously, if direct compression of the heart is the mechanism for forward movement of blood during chest compression, an arterial-venous pressure gradient across the heart must be produced during chest compression. The heart itself might serve as a pump without an arterial-venous pressure gradient across the heart if there was a marked phase difference between high arterial and high venous pressures. But in our dog studies there was neither a systolic pressure gradient across the heart nor a marked phase difference in the rise and fall of arterial and right atrial pressures.

In several of over 200 dogs we studied, a marked intrathoracic arterial-venous pressure gradient did exist in which aortic pressure was much higher than right atrial pressure with chest compression. The arterial pressure in these dogs exceeded 100 mm Hg during chest compression, which is far higher than usual for large dogs during conventional CPR. Thus, compression of the heart through sternal compression of even animals with a chest structure such as the dog can result in direct cardiac compression. In the absence of an intrathoracic arterial-venous pressure gradient, even if the heart is physically compressed, cardiac compression is not necessarily responsible for the peripheral or systemic blood flow (see Figure 1). The factors responsible for forward flow are those factors generating the peripheral arterial-venous gradient. If cardiac compression and a rise in intrathoracic pressure both occur, the effects will likely be additive.

Radionuclide angiographic studies charting the flow of blood through the chest (7) and angiograms performed during conventional CPR (18) support the early observations with regard to mechanism. The initial flow of blood following the rise in intrathoracic pressure is from the aorta, which partially collapses as blood is sucked toward the periphery from the aorta itself. Next, the valve structures of the heart open and blood moves from the lung and left atrium through the left ventricle into the aorta. There is little change in overall left ventricular size. This latter angiographic observation (18) strongly supports the notion that the heart is not compressed in the dog.

Therefore, during CPR the blood apparently circulates because the increased intrathoracic pressure is transmitted equally or nearly equally to all intrathoracic vascular structures. These intravascular pressures are unequally transmitted into the extrathoracic arterial and venous beds, which creates an extrathoracic arterial-venous pressure gradient and forward blood flow during the period of high thoracic pressure. During diastole (release of chest compression), blood flows into the lungs as a result of the extrathoracic venous-to-intrapulmonary pressure gradient. Compression of the right ventricle during chest compression may aid in pulmonary flow. Overall then, the blood moves to and from the lungs by virtue of the changes in intrathoracic pressure. The heart serves only as a conduit for the passive movement and conduct of blood to and from the lungs. The valves of the heart are not of major significance with regard to extrathoracic blood flow. The mitral and the aortic valves must be open during systole, and the tricuspid and pulmonary valves must be open during diastole. During systole the tricuspid and pulmonary valves may close (although there are no data in this regard) and may serve as ancillary valves in an additive fashion to the multiple venous valves in preventing retrograde flow. Our own suspicion is that the tricuspid valve, because of its large size and structure, is likely incompetent. Diastolic closure of the aortic valve may be critically important in maintaining coronary blood flow.

STUDIES IN MAN

Movement of Blood

We cannot yet draw any final conclusions as to the frequency or importance in man of the two mechanisms (chest compression or generalized increase in intrathoracic pressure) with conventional CPR.

In a number of patients we demonstrated (20) the presence of a pressure gradient at the thoracic inlet upon withdrawing intravascular catheters from the superior vena cava to the extrathoracic internal jugular vein. Since this pressure gradient does exist in man and since valves at this site can be demonstrated rather easily in man during cough (9), this newly identified mechanism for forward movement of blood is certainly operative. On the other hand some patients have readily measureable high arterial pressure with conventional CPR, pressures higher than those usually generated in man during conventional CPR and also higher than those in most dogs. It is our impression that most patients who have low radial artery pressure during CPR are probably able to move blood through a mechanism related solely to the rise in intrathoracic pressure.

It is not essential in the human to think about these mechanisms in an exclusive fashion. Direct cardiac compression is useful when possible; when it is not potent enough to maintain cerebral perfusion, manipulation of intrathoracic pressures would likely have a favorable additive effect on carotid blood flow (5).

It must be emphasized that, where effective, direct cardiac compression has one particularly attractive feature. With direct cardiac compression it is likely that the aortic diastolic pressure will be significantly higher than when blood is moved solely through manipulation of intrathoracic pressure. In terms of coronary perfusion, this would clearly be a useful feature.

IMPROVEMENT OF CPR

We have explored two basic strategies for improving peripheral perfusion during CPR. Both depend on the general principle of moving blood by raising intrathoracic pressure. The first strategy is to increase intrathoracic pressure per se. Increasing intrathoracic pressure through clamping of the endotracheal tube, while continuing chest compression augmented carotid blood flow in a manner paralleling the increase in intrathoracic vascular pressures (20). During these experiments in the dog it soon became apparent that one of the limitations of such manipulation of intrathoracic pressure was the occurrence of carotid arterial collapse. Carotid arterial collapse will occur when the intraarterial pressure falls below the sum of the extraarterial pressure (intrathoracic pressure) and the resistance of the artery to collapse. When carotid collapse occurs there can, in fact, be a reduction in flow through the increase in intrathoracic pressure.

With merely clamping the airway, there is a marked initial increase in extrathoracic carotid pressure and flow with the first chest compression. These benefits dissipate rapidly with subsequent compression cycles, undoubtedly because venous return is inhibited by the continuously high intrathoracic pressure following clamping of the airway. To take advantage of this mechanism for increase in blood flow without inhibiting venous return, the technique of high airway pressure ventilation synchronous with chest compression was developed. Such techniques were shown previously to improve hemodynamics and animal survival (22), but the mechanism was unclear and the results doubted (12, 14).

The second approach to augmenting blood flow during cardiopulmonary resuscitation is related to abdominal binding. Some years ago abdominal binding was proven useful in improving the hemodynamics of cardiopulmonary resuscitation (19). One large study in dogs reported significant benefits in survival and hemodynamics without any significant untoward complications. In contrast, another study seemed to show a higher incidence of liver rupture resulting from abdominal binding (11).

In the dog we were able to show significant augmentation of carotid blood flow through abdominal binding. Abdominal binding likely improves carotid blood flow for a number of reasons. First, with conventional CPR abdominal binding alone increases esophageal pressure and vascular pressures during chest compression, which may reflect a greater rise in intrathoracic pressure. Secondly, abdominal binding may direct blood flow away from organs below the diaphragm. Thirdly, abdominal binding increases the circulating blood volume (20).

SYNCHRONOUS VENTILATION AND CHEST COMPRESSION

Using carotid blood flow measured with a cannulating flow probe as the sole end-point for assessing efficacy, we compared conventional CPR (chest compression alone at a rate of 60/min with ventilation interposed after every fifth compression) with a new approach. This new approach includes high airway pressure ventilation synchronous with chest compression at variable rates. Venous return is allowed to occur by interrupting both ventilation and compression between each ventilation-compression cycle (diastole). Synchronous ventilation and chest compression alone approximately doubles carotid flow (3). There is an additive beneficial effect from abdominal binding and also from volume loading. It should be noted that in some dogs synchronous ventilation and compression provides a lower carotid blood flow than conventional CPR. In these animals the carotid artery collapses. Carotid collapse can be prevented or reversed in most dogs by abdominal binding, which likely increases the circulating blood volume. In addition, we have most recently found that negative diastolic airway pressure will reverse carotid collapse in some dogs (2), probably because pulmonary blood flow increases when the negative diastolic airway pressure supplies more blood for subsequent circulation.

We have applied abdominal binding (6) and ventilation at high airway pressures synchronous with chest compression (4) for brief periods to a small number of patients. Although the results of these studies are encouraging, such factors as adequacy of ventilation and the benefits or complications of these procedures remain to be explored.

Literature Cited

1. Brecher, G. A. 1952. Mechanism of venous flow under different degrees of aspiration. *Am. J. Physiol.* 169–423
2. Chandra, N., Cohen, J. M., Tsitlik, J., Weisfeldt, M. L. 1979. Negative airway pressure between compressions augments carotid flow during CPR. *Circulation* 59 & 60:II–46
3. Chandra, N., Rudikoff, M., Tsitlik, J., Weisfeldt, M. L. 1979. Augmentation of carotid flow during cardiopulmonary resuscitation (CPR) in the dog by simultaneous compression and ventilation with high airway pressure. *Am. J. Cardiol.* 43:422
4. Chandra, N., Rudikoff, M., Weisfeldt, M. L. 1980. Simultaneous chest compression and ventilation at high airway pressure during cardiopulmonary resuscitation. *Lancet* 1:175–78
5. Chandra, N., Snyder, L., Tsitlik, J., Weisfeldt, M. L. 1980. Non-invasive assisted circulation by synchronized cyclical high intrathoracic pressure support. *Circ. Res.* 28:161A
6. Chandra, N., Snyder, L., Weisfeldt, M. L. 1979. Abdominal binding during CPR in man. *Circulation* 59 & 60:II–45
7. Cohen, J. M., Alderson, P. O., Van Aswegen, A., Chandra, N., Tsitlik, J., Weisfeldt, M. L. 1979. Timing of intrathoracic blood flow during resuscitation with high intrathoracic pressure. *Circulation* 59 & 60:II–196
8. Criley, J. M., Blaufuss, A. N., Kissel, G. L. 1976. Cough-induced cardiac compression. *J. Am. Med. Assoc.* 236:1246–50
9. Fisher, J., Chandra, N., Eaton, L., Weisfeldt, M. L. 1980. Venous hemodynamics during cough in man. *Circ. Res.* 28:168A
10. Franklin, K. J. 1927. Valves in veins: an historical survey. In *Proc. R. Soc. Med., Sect. History Med.,* pp. 4–6
11. Harris, L. C., Kirimli, B., Safar, P. 1967. Augmentation of artificial circulation during cardiopulmonary resuscitation. *Anesthesiology* 28:730–34
12. Harris, L. C., Kirimli, B., Safar, P. 1967. Ventilation-cardiac compression rates and ratios in cardiopulmonary resuscitation. *Anesthesiology* 28:806–12
13. Holt, J. P. 1941. The collapse factor in the measurement of venous pressure. *Am. J. Physiol.* 134:292
14. Jude, J. R. 1963. Discussion. *Surgery* 53:193–94
15. Kouwenhoven, W. B., Jude, J. R., Knickerbocker, G. C. 1980. Closed heart cardiac massage. *J. Am. Med. Assoc.* 173:1064–67
16. MacKenzie, J. 1894. The venous and liver pulses, and the arhythmic (sic) contraction of the cardiac cavities. *J. Pathol. Bacteriol.* 2:113
17. MacKenzie, G. J., Taylor, S. H., McDonald, A. H., Donald, K. W. 1964. Hemodynamic effects of external cardiac compression. *Lancet* 1:1342–45
18. Niemann, J. T., Garner, D., Rosborough, J., Criley, J. M. 1979. The mechanism of blood flow in closed chest cardiopulmonary resuscitation. *Circulation* 59 & 60:II–74
19. Redding, J. S. 1971. Abdominal compression in cardiopulmonary resuscitation. *Anesthesia Analgesia* 50:668–75
20. Rudikoff, M. T., Maughan, W. L., Effron, M., Freund, P., Weisfeldt, M. L. 1980. Mechanisms of blood flow during cardiopulmonary resuscitation. *Circulation* 61:345–52
21. Weale, F. E., Rothwell-Jackson, R. L. 1962. The efficiency of cardiac massage. *Lancet* 1:990–92
22. Wilder, R. J., Weir, D., Rush, B. F., Ravitch, M. M. 1963. Method of coordinating ventilation and closed chest cardiac massage in the dog. *Surgery* 53:186–94
23. Yin, F. C. P., Cohen, J. M., Tsitlik, J., Weisfeldt, M. L. 1979. Arterial resistance to collapse: A determinant of peripheral flow resulting from high intrathoracic pressure. *Circulation* 59 & 60:II–196

Ann. Rev. Med. 1981. 32:443–56
Copyright © 1981 by Annual Reviews Inc. All rights reserved

PATHOGENESIS, DIAGNOSIS, AND MANAGEMENT OF REFLUX ESOPHAGITIS

❖7411

D. Kestenbaum, M.D. and J. Behar, M.D.

Department of Medicine, Division of Gastroenterology, Rhode Island Hospital
and Brown University School of Medicine, Providence, Rhode Island 02902

INTRODUCTION

The most common symptoms associated with reflux esophagitis are heartburn and regurgitation. Heartburn is usually described as a substernal or retrosternal burning pain or discomfort with a moving quality from the xyphoid area up toward the suprasternal notch. The pain may radiate up toward the neck and jaw as well as down into the epigastric area. Rarely, the pain may radiate into the arms and simulate the pain of cardiac origin (1). Heartburn is frequently associated with regurgitation of sour or bitter fluid. These symptoms are exacerbated by meals or postural changes, usually made worse when the patient is recumbent or tenses his abdominal muscles as in lifting. Most patients (84%) experience symptoms of reflux for more than one year before seeking medical attention (2).

Intermittent dysphagia occurs frequently with the first swallows of every meal. The symptom is often reported by patients with severe reflux esophagitis. They do not often have mechanical obstruction to the passage of food. Usually the dysphagia will be mild and nonprogressive with no associated weight loss. This is the result of a motility disturbance within the body of the esophagus induced by reflux of gastroduodenal contents (3) and occurs most frequently with the first bolus swallowed. Persistent dysphagia of solids in a patient with reflux esophagitis should indicate to the clinician that a peptic stricture has formed or that a carcinoma may have developed within the esophagus. Anemia has been described as the presenting symptom in some patients with reflux esophagitis; it is a result of low-grade

0066-4219/81/0401-0443$01.00

bleeding from erosions and ulcerations in the esophagus. The incidence of this complication is about 5% (4). Major gastrointestinal bleeding requiring blood transfusions, however, is uncommon. Major bleeding is usually preceded by the ingestion of exogenous caustic agents, gastric intubation, or retching and vomiting.

There is increasing evidence that aspiration of gastroduodenal contents in the respiratory tract and lungs can occur in patients with reflux esophagitis. Morning hoarseness, asthma-like symptoms, and aspiration pneumonitis have been reported in these patients. Pulmonary aspiration of gastroduodenal contents appears to be more common in the pediatric population (5–7). Lung scanning with the gamma camera using technetium as radioisotope may simplify the diagnosis of this complication, and its frequency and importance may become clearer with time (8).

PATHOGENESIS

Abnormal Anti-Reflux Barrier

The primary motor abnormality in this disease entity is the weakness of the lower esophageal sphincter (LES) and its failure to prevent gastroesophageal reflux (GER) (9). Approximately 75% of patients with reflux esophagitis have basal LES pressures of 12 mm Hg or less compared to normal subjects with pressures higher than 12 mm Hg (4). The pathogenesis of the sphincter incompetence (see Figure 1) is poorly understood since we know little about the mechanisms that control the LES in man. Most of our understanding of the sphincter comes from experimental studies in animals and from clinical models such as human achalasia (10). A competent LES prevents GER by maintaining a high-pressure zone at the gastroesophageal junction in the resting state. Experimental studies suggest that three factors —the circular muscle of the sphincter, the innervation of the gastroesophageal junction, and gastrointestinal neuropeptides—may contribute to the competence of the sphincter. These may interact in a complex and not completely understood way in determining the resting and adaptive pressures of the LES.

The circular muscle layer of this region is considerably specialized. In vivo studies show that the LES maintains a resting high-pressure zone that is either not affected at all (opossum) or only slightly reduced (cat) by complete denervation with neural poisons such as tetrodotoxin (11, 12). These studies indicate that basal LES pressures are determined by the muscle itself. In vitro studies that the circular muscle strips from this region develop an active tension greater than adjacent strips of circular muscles from the esophagus or fundus and are not affected by complete denervation by tetrodotoxin (13, 14). These findings cannot be entirely explained by the

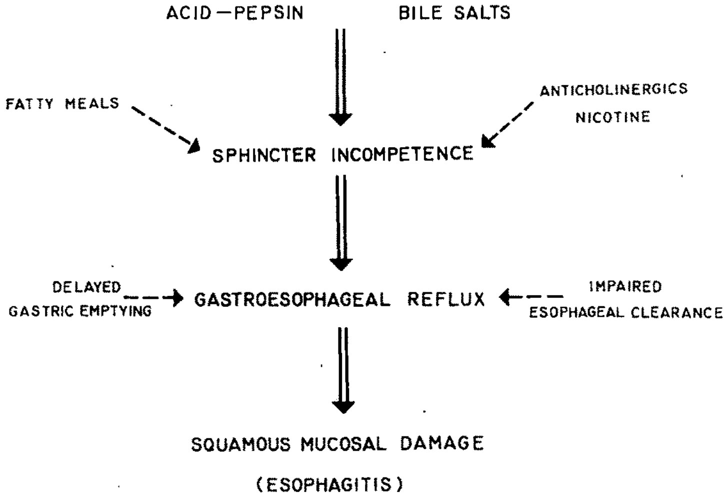


Figure 1 Pathogenesis of gastroesophageal reflux and esophagitis.

greater muscle thickness of this region. The myogenic mechanisms of this specialized circular muscle layer may be related to lower membrane potentials allowing an influx of calcium in the LES muscle cell (15). Furthermore, muscle cells from the LES, but not from the body of the esophagus, contain receptors for prostaglandins (16). LES strips are also capable of converting arachidonic acid, a prostaglandin precursor, into contractile substances. This conversion can be prevented by agents that interfere with prostaglandin synthesis, for example indomethacin, which causes a decrease in LES active muscle tension. It is thus conceivable that endogenous prostaglandin may maintain a high resting tonic contraction of the circular muscle layer of the LES by acting locally at specific receptors in the cell membrane that could facilitate the influx of calcium into these cells (17).

The neural contribution to the genesis of the LES tonic contraction varies considerably among species. In the opossum, neural regulation may be due to a balance between excitatory and inhibitory influences, since sphincter denervation with tetrodotoxin results in no discernable pressure change (11). In the cat, however, it appears that an excitatory alpha-adrenergic influence predominates since both tetrodotoxin and phentolamine, an alpha-adrenergic blocker, decrease LES pressures by about 20–30% (12). In man and in dogs, cholinergic influences appear to play a significant excitatory role in the genesis of LES tonic contraction. Small doses of atropine reduce LES pressures by about 40% in man (18, 19) and almost completely in dogs (20).

The role of gastrointestinal neuropeptides in the control of the LES competence is even more controversial. Their potential role has been suggested by the following observations: (*a*) the presence of receptors for these neuropeptides in mediating either LES contraction or relaxation (21–24); (*b*) the presence of these neuropeptides at the autonomic nerve endings that innervate the esophagus (25); and, (*c*) the significant changes in basal LES pressures during the digestive process when these neuropeptides are released into circulation (26). The correlation between LES pressures and circulating levels of these hormones, however, is poor (27, 28). The neuropeptides may conceivably act at local levels as neurotransmitters or as paracrine factors modulating synaptic transmission (29). A great deal of additional work is needed to define whether or not there is a specific role for these hormones in the control of LES tonic contraction. Nevertheless, marked elevation of the circulating levels of some of these hormones in physiological and diseased states may cause changes in the strength of the LES circular muscle. An example of a specific hormonal change causing LES incompetence and GER is observed during pregnancy. Reflux symptoms are common in pregnancy and can be very severe. In the majority of patients, the symptoms resolve after delivery (30). It has been suggested that the LES incompetence is induced by high circulating levels of progesterone and estrogens. These assumptions are supported by studies performed in young women taking oral contraceptives (31) and in experimental animals with the pseudopregnancy induced by the administration of female hormones (32). Thus, we can conclude that the sphincter incompetence in patients with reflux esophagitis may result from abnormalities in either the muscle, neural fibers, or gastrointestinal neuropeptides singly or in combination.

Esophageal manometry has disclosed that in about 25% of patients with reflux esophagitis LES pressures overlap with those of normal subjects (4). One possible explanation is that basal LES pressures, as measured under laboratory conditions, may not always reflect those pressures present when GER develops or that the adaptive mechanisms of the LES may still be abnormal in these patients. This LES adaptation occurs in response to an increase in intraabdominal or intragastric pressures (33). There has been considerable controversy about the existence and significance of this LES pressure response. Measurement of this adaptive response in man has been clouded by technical difficulties since it is difficult to maintain the pressure sensor in the same site within the LES before and during these responses. Initially, it was suggested that this was the result of a neural reflex. More careful human and animal studies, however, seem to suggest that this pressure change may result when the LES is mechanically compressed by the diaphragm or other anatomic structures (34, 35).

Alternatively, it might be that other mechanisms participate in the control of GER. It has been assumed that a sliding hiatus hernia interferes with or eliminates the action of mechanical factors that assist the LES in the control of GER. Thus the absence of these mechanical factors may make reflux possible even though the LES is normal. The most important mechanical factors appear to be the mucosal flap valve, the acute esophageal angle of Hiss, the compression of the distal esophagus by diaphragmatic hiatus, the interdigitation of the esophagogastric folds (the mucosal choke), the phrenoesophageal membrane, and, perhaps most importantly, the intraabdominal LES extrinsically compressed by the positive pressures of this cavity. The role of these mechanical factors, however, has not stood the test of careful studies in dog models. The creation of a hiatus hernia in dogs, for instance, does not result in esophagitis nor does it change the resting LES pressures (36). Although there is an undeniable strong association between the presence of a hiatus hernia and reflux esophagitis, it is not a simple one. Cohen & Harris (37) found that reflux symptoms correlated well with the presence of an incompetent LES but not with the presence or absence of a hiatus hernia. Furthermore, this association does not necessarily imply a cause-to-effect relationship. The exact incidence of hiatus hernia in the general population is not clearly known. There are difficulties in the diagnosis with both radiological and endoscopic techniques. Some studies suggest that up to 50% of subjects that undergo an upper gastrointestinal x-ray examination have a sliding hiatus hernia, and the incidence frequently depends on the radiologist's efforts to demonstrate this hernia (38). In a few patients, the hiatus hernia is large with a wide diaphragmatic hiatus and the phrenoesophageal membrane is either ruptured or elongated. Evidence of reflux esophagitis is rarely found in this type of patient. A hiatus hernia is also observed in patients with severe reflux esophagitis and occasional stricture formation. It is conceivable that in these patients the esophagus is shortened as the result of a chronic inflammation. Not only does the esophagus contract radially forming a stricture but longitudinally as well, thereby pulling the stomach up into the chest.

In addition there are mechanisms that aggravate reflux once it occurs. The normal esophagus is able to clear the refluxed contents through its primary and secondary peristalsis. This acid clearance function enables the esophagus to reduce the duration and damaging effect of the GER. Patients with reflux esophagitis often have a decreased esophageal acid clearance when compared to normal subjects (39). When reflux occurs in patients with severe GER, it tends to last longer, particularly in the recumbent position (40). There is some evidence that the acid clearance function of the esophagus is impaired during sleep. More recently it has been shown that even normal subjects can reflux in their sleep because of spontaneous or inappropriate relaxations of the LES (41). It is possible that in some patients with reflux esophagitis, but with normal basal LES pressures, severe GER may occur because of frequent spontaneous or inappropriate relaxations of the sphincter and an associated impaired esophageal acid clearing mechanism. We have found abnormal esophageal peristalsis in patients with severe reflux esophagitis with contractions of low amplitude or force, almost mimicking the manometric findings observed in patients with Raynaud's phenomenon or scleroderma (J. Behar, unpublished observations). A few patients also exhibit a very disorganized motor activity of the esophagus that simulates moderate to severe diffuse esophageal spasm. It is not clear, however, whether these motor disorders are part of the primary process of this disease or secondary changes due to damage caused to the neuromuscular function of the esophagus by GER.

Consequences of Gastroesophageal Reflux

Reflux of gastroduodenal contents into the esophagus damages the squamous mucosa; the degree of damage is related to the pH and pepsin activity and the potentiating effects of bile salts (42, 43). This damage can be detected by changes in the potential difference across the squamous mucosa. The initial mucosal damage is apparently an increase in surface epithelial desquamation and compensatory cellular proliferation. This is reflected morphologically by an elongated papilla and thicker basal layer, which is the proliferative cellular compartment (44, 45). This greater cell proliferation has been suggested because the uptake in tritiated thymidine by the esophageal mucosa increases in vitro, as judged by biopsies from patients with severe esophagitis and from experimental esophagitis in cats (46, 47). Persistence of gastroesophageal reflux may lead to mucosal ulcerations when the cell proliferation process fails to compensate fully for the increased cellular desquamation. Experimental studies in dogs show that if the mucosal ulcerations are continuously bathed by the refluxed contents, they tend to heal, partially or completely, with a metaplastic mixed epithelium (Barrett's epithelium) (48), which is characterized by a combination of gastric, fundic, pyloric, or intestinal glands (49). There is increasing evidence that this metaplastic epithelium is premalignant predisposing to development of adenocarcinoma (50, 51). In contrast, when the GER is controlled or prevented, the ulcerations tend to heal with a normal squamous epithelium. A recent study revealed a dynamic interaction between these metaplastic and normal squamous epithelia. Even after the metaplastic epithelium is well established, controlling the GER by surgical means can lead to its regression and replacement by normal squamous epithelium (52).

The inflammatory response is not confined to the squamous mucosa but can extend to the submucosa, muscularis propria, and periesophageal tissues. Inflammation of the esophageal wall and periesophageal tissues can induce a marked fibrotic response with formation of a stricture that interferes with the passage of a bolus of food. Furthermore, damage by this inflammatory process to the LES circular muscle layer can lead to further weakening of the sphincter, as shown in experimental esophagitis in cats in vivo and in vitro (53, 54).

DIAGNOSIS

Based on the pathogenesis of this disorder, patients with reflux esophagitis have an incompetent LES that causes GER and subsequent injury on the esophageal squamous mucosa or esophagitis. It seems logical to base the diagnosis of this disease on objective evidence of these two distinct abnormalities (4).

Assessing Sphincter Incompetence

Esophageal manometry allows us to determine basal LES pressures as well as esophageal peristalsis in response to swallowing. Most patients with reflux esophagitis have basal LES pressures of 12 mm Hg or less, whereas normal subjects have pressures greater than 12 mm Hg. The remaining 25% of patients with reflux esophagitis have LES pressures that overlap with normal subjects. In these patients, sphincter incompetence can be detected by demonstrating GER of acid using a minute pH probe placed 5 cm above the LES (9). This test can demonstrate GER in about 90% of patients with reflux esophagitis, whereas only 10% of normal subjects have false positive tests. GER can also be detected and semiquantitated by using the scintiscanning technique with the radioisotope 99-technetium and the gamma camera (55). Also, it has been suggested that the diagnosis of sphincter incompetence be accomplished by monitoring the esophagus for GER of acid with a pH electrode for periods of 12–24 hours. Although this technique is extremely useful as a research tool, it does not seem to have any diagnostic importance at this time since it has not been quantitated.

Determining Esophagitis

Esophageal infusion of an acid solution (HCl 0.1 N) or Bernstein test (56) is a test designed to elicit the patient's substernal discomfort. It is only confirmatory in patients who have classical and well-defined heartburn but it can be useful when there is a doubt that the esophagus is the source of a substernal discomfort and can be determined by eliciting the patient's pain during the acid infusion. However, it does not necessarily indicate the presence of esophagitis since patients with "a sensitive esophagus" and diffuse esophageal spasm may also have a positive test. A negative test result also is useful in that it rules out esophagitis in the majority of patients. We found that only 12% of patients who had the diagnosis of reflux esophagitis had a negative test (4). Esophagoscopy is very specific but not very sensitive since it only discloses muscosal lesions in about 60% of the patients (4). The remaining 40% have a normal esophageal mucosa at esophagoscopy. It allows assessment of the severity of the disease since patients with gross esophagitis with mucosal ulcerations can develop complications such as stricture, low-grade bleeding, and Barrett's epithelium. It also allows direct visual biopsies to detect the presence of Barrett's epithelium. The biopsy of the esophagus, using suction techniques, should be performed in patients who have a normal esophagoscopy; such biopsies usually do not reveal any evidence of acute inflammatory reaction. They do reveal suggestive epithelial changes, such as an elongated papilla and the thicker basal layer, that can be detected in the majority of these patients (44, 45).

MANAGEMENT

In managing patients with reflux esophagitis, we should keep in mind four considerations: (*a*) we are dealing with a chronic disease that requires some form of continuous treatment; (*b*) the esophagus is at higher risk when these patients sleep in the recumbent position because the gravity factor is absent and esophageal acid clearance is impaired during sleep; (*c*) the severe inflammatory process might damage the neuromuscular structures of the esophagus, worsen the sphincter incompetence, and perhaps impair the esophageal acid clearance; and, (*d*) gross esophagitis can lead to complications that to a certain degree, may be reversible.

Medical Management

The goals of medical management are to improve sphincter competence and neutralize or inhibit the secretion of causative agents such as acid, pepsin, and bile salts.

Drugs or measures that improve sphincter competence also seem to facilitate the esophageal clearance of GER and lessen its damaging effects. These goals can be obtained by elevating the head of the bed 6–8 inches, which restores the gravity factor at the gastroesophageal junction while the patient is asleep. The duration of GER is reduced by this maneuver (57). Dietary manipulations can also be helpful since protein with meals increases LES pressures whereas fatty meals, alcohol, and nicotine affect the sphincter competence adversely (58). Certain drugs can also be harmful, particularly anticholinergics, which not only reduce LES pressures but also delay gastric emptying.

We can also use drugs that improve the LES competence and increase the magnitude or force of esophageal contractions. At the present time, there are two drugs that can accomplish these two objectives. One is bethanechol (Urecholine®), which is used in the oral dose of 25 mg, 3–4 times a day. In a double-blind crossover study it improved reflux symptoms and decreased the consumption of antacids when compared to placebo (59). Its efficacy, however, is limited where it is needed the most, in patients with moderate to severe esophagitis. Bethanechol also has some drawbacks; a few patients complain of abdominal cramps and constipation or diarrhea, and there is some concern over its long-term effects on gastric acid secretion. The other drug is metoclopramide, which improves sphincter competence and reduces the duration of GER. Metoclopramide in doses of 10 mg, 3–4 times a day, does not cause abdominal cramps nor does it stimulate gastric acid secretion. Although it has been used in Europe for a number of years, only one controlled clinical trial showed that this drug causes symptomatic relief and increased esophageal tolerance to acid when compared to placebo and moderate antacid ingestion (60). Its major disadvantage results from its effects on the central nervous system. Patients frequently complain of increased anxiety and restlessness and occasionally even agitation. Nevertheless, either drug may find an important place in the therapy of this disease when used in combination with other forms of therapy.

For years, reflux esophagitis patients have been treated with antacids immediately or one hour after their meals. Antacids are very effective in relieving the symptoms of heartburn but they are less successful in preventing this reflux symptom for a long period of time or in changing the natural course of the disease. In a prospective study of antacids used four times a day, only 20% of patients were able to be symptom free (61). The remaining 80% complained of intermittent heartburn with 20% of these patients developing complications such as early stricture and low-grade GI bleeding. H-2 receptor blockers such as cimetidine appear to be effective in patients with reflux esophagitis. Recent studies on both sides of the Atlantic have shown that cimetidine at doses of 300 mg, 4 times a day, is more effective than placebo in relieving the frequency and severity of the heartburn, particularly at night, and in improving the tolerance of the esophagus to acid infusion (62). However, within the 8-week treatment period, the endoscopic appearance did not improve.

Although these newer drugs appear to be very promising, we need to determine whether or not they can change the natural history of the disease over a longer period of time. Furthermore, the efficacy of treatments with combinations of drugs that improve sphincter competence and that reduce acid secretion is not known. Such clinical trials would be extremely useful indeed.

Surgical Management

Patients with reflux esophagitis who fail to respond to an adequate medical trial should be considered candidates for surgical management. Three groups of patients should be considered for this type of management. The first are patients with intractable reflux symptoms. The second group are those who develop complications of gastroesophageal reflux disease during an adequate medical trial. These complications include formation of an early stricture and persistent or recurrent low-grade bleeding leading to anemia. Another complication, the presence of Barrett's epithelium in the esophagus, may also contribute to the decision to operate on a patient, since surgical management has been shown to cause its regression and replacement by a normal squamous epithelium (52). The third group of patients to be considered for surgery are those with objective and preferably direct evidence of pulmonary aspiration caused by gastroesophageal reflux disease (8). Most of the previous studies have shown indirect evidence of this pulmonary aspiration since direct evidence has so far proved elusive.

Surgical treatment is effective in controlling severe gastroesophageal reflux and its consequences (61, 63, 64). Hiatal hernia repairs with a fundoplication of Nissen, anterior fundoplication of Belsey, or posterior gastropexy of Hill performed in patients with reflux esophagitis lead to symptomatic improvement and concomitant resolution of the objective changes of esophagitis. After this surgical manipulation, there is conversion of the positive to a negative acid infusion test, healing of gross esophagitis at esophagoscopy, and prevention of acid reflux.

In considering a patient with reflux esophagitis for surgical treatment, one should keep in mind that two major questions about this type of management remain unanswered (65). The first one of great practical importance is the effective duration of these surgical manipulations. There is only one careful but uncontrolled study showing that the incidence of recurrence of reflux esophagitis increases with time (66). It is not clear why symptoms recur, but it has been assumed that it results from the breakdown of the repair and recurrence of esophageal hiatal hernia. A more recent study, however, has shown that most patients with recurrence of reflux symptoms have an intact anatomic repair at least when examined radiologically (67). This study would suggest that recurrence of GER may be due to progression of the primary disorder responsible for LES incompetence.

The question of the effectiveness of the repair and its duration is intimately related to the second question of how the surgical reconstruction of the cardioesophageal junction controls GER. There is now increasing evidence that these repairs not only reconstruct the anatomy of the gastroesophageal junction but also increase the LES strength. Several studies have shown that hiatal hernia repairs increase LES pressures immediately and at least up to three years postoperatively (61, 68). The mechanism of the rise in sphincter pressures, however, is far from known. Surgical manipulation of the gastroesophageal region could make the sphincter more competent by creating a flap valve which would facilitate the transmission of changing gastric pressures to the LES. Greater LES pressures postoperatively have also been explained by the restitution of the sphincter to its normal subdiaphragmatic position, where it could then take advantage of higher intraabdominal pressures. The postoperative sphincter pressures, however, are greater than the simple sum of the preoperative sphincter and postoperative gastric pressures (69). Furthermore, sphincter pressures are not affected by displacement of the normal canine LES into the chest. Over the last few years two hypotheses have been advanced to explain this phenomenon of postoperative rise in LES pressures (70, 71). The LES may be buttressed by the fundal cuff placed around the distal esophagus. Fundal muscles share some of the physiologic and pharmacologic characteristics of LES muscles. They relax during swallowing and vagal stimulation and contract in response to pentagastrin. It is thus possible that this fundal cuff strengthens the LES without altering its physiological characteristics. It is also conceivable that surgical manipulation of the distal esophagus might alter the length-tension characteristics of the LES circular muscle. Fundoplication may thicken the mucosal folds or shorten the circular muscle layer, which could then stretch the muscle fibers with greater tension at closure.

ACKNOWLEDGMENTS

The authors are grateful to Lorie Klemer and Carol Marin for their assistance in the preparation of this manuscript.

Literature Cited

1. Bennett, J. R., Atkinson, M. 1966. Oesophageal acid-perfusion in the diagnosis of precordial pain. *Lancet* 2: 1150–52
2. Brunnen, P. L., Karmody, A. M., Needham, C. D. 1969. Severe peptic esophagitis. *Gut* 10:831–37
3. Olsen, A. M., Schlegel, J. F. 1965. Motility disturbances caused by esophagitis. *J. Thorac. Cardiovasc. Surg.* 50:607–12
4. Behar, J., Biancani, P., Sheahan, D. G. 1976. Evaluation of esophageal tests in the diagnosis of reflux esophagitis. *Gastroenterology* 71:9–15
5. Euler, A. R., Byrne, W. J., Ament, M. E., Fonkalsrud, E. W., Strobel, C. T., Siegel, S. C., Katz, R. M., Rachelefesky, G. S. 1979. Recurrent pulmonary disease in children: A complication of gastroesophageal reflux. *Pediatrics* 63:47–51

6. Christie, D. L., O'Grady, L. R., Mack, D. V. 1978. Incompetent lower esophageal sphincter and gastroesophageal reflux in recurrent acute pulmonary disease of infancy and childhood. *J. Pediatr.* 93:23–27
7. Kennedy, J. H. 1962. "Silent" gastroesophageal reflux: an important but little known cause of pulmonary complications. *Dis. Chest* 42:42–45
8. Chernow, B., Johnson, L. F., Janowitz, W. R., Castell, D. O. 1979. Pulmonary aspirations as a consequence of gastroesophageal reflux: a diagnostic approach. *Dig. Dis. Sci.* 24:839–44
9. Haddad, J. K. 1970. Relation of gastroesophageal reflux to yield sphincter pressures. *Gastroenterology* 58:175–84
10. Smith, B. 1970. The neurological lesion in achalasia of the cardia. *Gut* 11: 388–91
11. Goyal, R. K., Rattan, S. 1976. Genesis of basal sphincter pressure: effect of tetrodotoxin on lower esophageal sphincter pressure in opossum in vivo. *Gastroenterology* 71:62–67
12. Behar, J., Kerstein, M., Biancani, P. 1977. Neural control of lower esophageal sphincter (LES) closure. *Gastroenterology* 72:1029 (Abstr.)
13. Christensen, J., Conklin, J. L., Freeman, B. 1973. Physiologic specialization at the esophagogastric junction in three species. *Am. J. Physiol.* 225: 1265–70
14. Biancani, P., Zabinski, M. P., Kerstein, M. D., Behar, J. 1980. Lower esophageal sphincter mechanics: Anatomic and physiologic relationships of the esophago-gastric junction of cat. Submitted for publication
15. Fox, J. E., Daniel, E. E. 1978. The role of calcium in the genesis of lower esophageal sphincter (LES) tone and contraction. *Gastroenterology* 74:1035 (Abstr.)
16. Daniel, E. E., Sarna, S., Crankshaw, J. 1978. Mechanisms of tetrodotoxin-insensitive relaxation of opossum lower esophageal sphincter. *Proc. 6th Int. Symp. Gastrointestinal Motility, Edinburgh, 1977,* ed. H. L. Duthie, pp. 525–34. MTP Press
17. Daniel, E. E., Crankshaw, J., Sarna, S. 1978. Myogenic control of esophageal motor function. *2nd Int. Symp. on Esophagus and Gastroesophageal Junction, Ixtapa, Mexico,* 2:14–19
18. Behar, J., Kastendieck, J. 1974. Studies on sphincter competence. *Gastroenterology* 66:834 (Abstr.)
19. Dodds, W. J., Dent, J., Hogan, W. J., Arndorfer, R. C. 1978. The effect of atropine on esophageal motor function in man. *Gastroenterology* 74:1028 (Abstr.)
20. Lind, J. F., Crispin, J. S., McIver, D. K. 1968. The effect of atropine on the gastroesophageal sphincter. *J. Physiol. Pharmacol.* 46:233–38
21. Zwick, R., Bowes, K. L., Daniel, E. E., Sarna, S. K. 1976. Mechanism of action of pentagastrin on the lower esophageal sphincter. *J. Clin. Invest.* 57:1644–51
22. Jennewein, H. M. 1978. Motilin and lower esophageal sphincter. In *Gut Hormones,* ed. S. R. Bloom, 59:359–62. New York: Churchill Livingstone. 664 pp.
23. Rattan, S., Said, S. I., Goyal, R. K. 1977. Effect of vasoactive intestinal polypeptide (VIP) on the lower esophageal sphincter pressure (LESP). *Proc. Soc. Exp. Biol. Med.* 155:40–43
24. Behar, J., Field, S., Marin, C. 1979. Effect of glucagon, secretin, and vasoactive intestinal polypeptide on the feline lower esophageal sphincter: mechanisms of action. *Gastroenterology* 77:1001–7
25. Uddman, R., Alumets, J., Hakanson, R., Sundler, F., Walles, B. 1980. Peptidergic (enkephalin) innervation of the mammalian esophagus. *Gastroenterology* 78:732–37
26. Csendes, A., Oster, M., Brandsborg, O., Moller, J., Brandsborg, M., Amdrup, E. 1978. Gastroesophageal sphincter pressure and serum gastrin: Reaction to food stimulation in normal subjects and in patients with gastric or duodenal ulcer. *Scand. J. Gastroenterol.* 13:363–68
27. Henderson, J. M., Lidgard, G., Osborne, D. H., Carter, D. C., Heading, R. C. 1978. Lower oesophageal sphincter pressure response to gastrin—pharmacological or physiological? *Gut* 19:99–102
28. McCallum, R. W., Walsh, J. H. 1979. Relationship between lower esophageal sphincter pressure and serum gastrin concentration in Zollinger-Ellison Syndrome and other clinical settings. *Gastroenterology* 76:76–81
29. Grossman, M. I. 1979. Chemical messengers: a view from the gut. *Fed. Proc.* 38:2341–43
30. Lind, J. F., Smith, A. M., McIver, D. K., Coopland, A. T., Crispin, J. S. 1968. Heartburn in pregnancy—a manometric study. *Can. Med. Assoc. J.* 98: 571–74

31. van Thiel, D. H., Gavaler, J. S., Stremple, J. 1976. Lower esophageal sphincter pressure in women using sequential oral contraceptives. *Gastroenterology* 71:232–34
32. Shulze, K., Christensen, J. 1977. Lower sphincter of the opossum esophagus in pseudo-pregnancy. *Gastroenterology* 73:1082–85
33. Lind, J. F., Warrian, W. G., Wankling, W. J. 1966. Responses of the gastroesophageal junctional zone to increases in abdominal pressure. *Can. J. Surg.* 9:32–38
34. Dodds, W. J., Miller, W. N., Hogan, W. J., Arndorfer, R. C., Lydon, S. B., Stef, J. J., 1975. Influence of extrinsic forces on intraluminal lower esophageal sphincter pressure. *Gastroenterology* 68:885 (Abstr.)
35. Dodds, W. J., Hogan, W. J., Miller, Stef, J. J., Arndorfer, R. C., Lydon, S. B. 1975. Effects of increased intra-abdominal pressure on lower esophageal sphincter pressure. *Am. J. Dig. Dis.* 20:298–308
36. Sicular, A., Cohen, B., Zimmerman, A., Kark, A. E., 1967. The significance of an intra-abdominal segment of canine esophagus as a component antireflux mechanism. *Surgery* 61:784–90
37. Cohen, S., Harris, L. D. 1971. Does hiatus hernia affect competence of the gastroesophageal sphincter? *N. Engl. J. Med.* 284:1053–56
38. Wolf, B. S. 1973. Sliding hiatus hernia: The need for redefinition. *Am. J. Roentgenol. Radium Ther. Nucl. Med.* 117: 231–47
39. Stanciu, C., Bennett, J. R. 1974. Oesophageal acid clearing: one factor in the production of reflux esophagitis. *Gut* 15:852–57
40. Orr, W. C., Robinson, M. G., Johnson, L. F. 1979. Acid clearing during sleep in patients with esophagitis and controls. *Gastroenterology* 76:1213 (Abstr.)
41. Dent, J., Dodds, W. J., Friedman, R. H., Sekiguchi, T., Hogan, W. J., Arndorfer, R. C., Petrie, D. J. 1980. Mechanism of gastroesophageal reflux in recumbent asymptomatic human subjects. *J. Clin. Invest.* 65:256–67
42. Goldberg, H. I., Dodds, W. J., Gee, S., Montgomery, C., Zboralske, F. F. 1969. Role of acid and pepsin in acute experimental esophagitis. *Gastroenterology* 56:223–30
43. Henderson, R. D., Mugashe, F. L., Jeejeebhoy, K. N., Szczpanski, M. M., Cullen, J., Marryatt, G., Boszko, A. 1973. Synergism of acid and bile salts in the production of experimental esophagitis. *Can. J. Surg.* 16:12–17
44. Ismail-Beigi, F., Horton, P. F., Pope, C. E. II. 1970. Histological consequences of gastroesophageal reflux in man. *Gastroenterology* 58:163–74
45. Behar, J., Sheahan, D. G. 1975. Histological abnormalities in reflux esophagitis. *Arch. Pathol.* 99:387–91
46. Livstone, E., Sheahan, D. G., Behar, J. 1977. Studies of esophageal epithelial proliferation in patients with reflux esophagitis. *Gastroenterology* 73:1315–19
47. Livstone, E., Sheahan, D. G., Reinprecht, C., Contino, C., Selling, J., Biancani, P., McCallum, R. 1979. In vitro ^{3}H-thymidine uptake by esophageal mucosa in cat experimental esophagitis. *Gastroenterology* 76:1187 (Abstr.)
48. Bremner, C. G., Lynch, V. P., Ellis, F. H. Jr. 1970. Barrett's esophagus: congenital or acquired? An experimental study of esophageal mucosal regeneration in the dog. *Surgery* 68:209–16
49. Paull, A., Trier, J. S., Dalton, M. D., Camp, R. C., Loeb, P., Goyal, R. K. 1976. The histologic spectrum of Barrett's esophagus. N. Engl. J. Med. 295:476–80
50. Haggitt, R. C., Tryzelaar, J., Ellis, F. H., Colcher, H. 1978. Adenocarcinoma complicating columnar epithelium-lined (Barrett's) esophagus. *Am. J. Clin. Pathol.* 70:1–5
51. Berenson, M., Riddell, R. H., Skinner, D. B., Freston, J. W. 1978. Malignant transformation of esophageal columnar epithelium. *Cancer* 41:554–62
52. Brand, D. L., Ylvisaker, J. T., Gelfand, M., Pope, C. E. II. 1980. Regression of columnar esophageal (Barrett's) epithelium after anti-reflux surgery. *N. Engl. J. Med.* 302:844–48
53. Eastwood, G. L., Castell, D. O., Higgs, R. H. 1975. Experimental esophagitis in cats impairs lower esophageal sphincter pressure. *Gastroenterology* 69:146–53
54. Biancani, P., Dodds, W. J., Storer, E., Selling, J., McCallum, R. W. 1979. Acute experimental esophagitis affects mechanical properties of the lower esophageal sphincter. *Gastroenterology* 76:1100 (Abstr.)
55. Fisher, R. S., Malmud, L. D., Roberts, G. S., Lobis, I. F. 1976. Gastroesophageal (GE) scintiscanning to detect and quantitate GE reflux. *Gastroenterology* 70:301–8
56. Bernstein, L. M., Baker, L. A. 1958. A clinical test for esophagitis. *Gastroenterology* 34:760–81

57. Babka, J. C., Hager, G. W., Castell, D. O. 1973. The effect of body position on lower esophageal sphincter pressure. *Am. J. Dig. Dis.* 18:441–42
58. Nebel, O. T., Castell, D. O. 1972. Lower esophageal sphincter pressure changes after food ingestion. *Gastroenterology* 63:778–83
59. Farrell, R. L., Roling, G. T., Castell, D. O. 1974. Cholinergic therapy of chronic heartburn. A controlled trial. *Ann. Intern. Med.* 80:573–76
60. Winnan, J., Avella, J., Callachan, C., McCallum, R. W. 1980. Double-blind trial of metoclopramide versus placebo-antacid in symptomatic gastroesophageal reflux. *Gastroenterology* 78: 1292 (Abstr.)
61. Behar, J., Sheahan, D. G., Biancani, P., Spiro, H. M., Storer, E. H. 1975. Medical and surgical management of reflux esophagitis: a 38-month report on a prospective clinical trial. *N. Engl. J. Med.* 293:263–68
62. Behar, J., Brand, D. L., Brown, F. C., Castell, D. O., Cohen, S., Crossley, R. J., Pope, C. E. II., Winans, C. S. 1978. Cimetidine in the treatment of symptomatic gastroesophageal reflux: a double-blind controlled trial. *Gastroenterology* 74:441–48
63. Skinner, D. B., Belsey, R. H. R., Russell, P. S. 1967. Surgical management of esophageal reflux and hiatus hernia. *J. Thorac. Cardiovasc. Surg.* 53:33–54
64. Hill, L. D. 1967. An effective operation for hiatal hernia. An eight year appraisal. *Ann. Surg.* 106:681–92
65. Behar, J. 1979. Surgical treatment of reflux esophagitis: how well does it work? *Gastroenterology* 77:183–84 (Editorial)
66. Brand, D. L., Eastwood, I. R., Martin, D., Carter, W. B., Pope, C. E. II. 1979. Esophageal symptoms, manometry and histology before and after antireflux surgery. A long-term follow-up study. *Gastroenterology* 76:1393–1401
67. DeMeester, T. R., Johnson, L. F., Kent, A. H. 1974. Evaluation of current operations for the prevention of gastroesophageal reflux. *Ann. Surg.* 180:511–25
68. Csendes, A., Larrain, A. 1972. Effect of posterior gastropexy on gastroesophageal sphincter pressure and symptomatic reflux in patients with hiatal hernia. *Gastroenterology* 63:19–24
69. Behar, J., Biancani, P., Spiro, H. M., Storer, E. H. 1974. Effect of anterior fundoplication on lower esophageal sphincter competence. *Gastroenterology* 67:209–15
70. Siewert, R., Jennewein, H. M., Waldeck, F. 1973. Mechanism of action of fundoplication. In *Proc. 4th Int. Symp. Gastrointestinal Motility,* ed. E. E. Daniel, p. 144–52. Alberta, Canada
71. Biancani, P., Zabinski, M., Behar, J. 1975. Pressure, tension, and force of closure of the lower esophageal sphincter and esophagus. *J. Clin. Invest.* 58:476–83

Ann. Rev. Med. 1981. 32:457–73
Copyright © 1981 by Annual Reviews Inc. All rights reserved

THE REGULATION OF RENAL CALCIUM EXCRETION: An Approach to Hypercalciuria

❖7412

Andrew F. Stewart, M. D., and Arthur E. Broadus, M.D.

Yale University School of Medicine, LMP 2074, 333 Cedar Street, New Haven, Connecticut 06510

INTRODUCTION

Urinary calculi are a major cause of morbidity in the United States. It is estimated that one of every 1000 people in the US is hospitalized annually for urinary tract stones, and that 15 of every 100 stone episodes results in urologic surgery (1).

Renal stones are classified into five major crystalline types: cystine, uric acid, struvite (magnesium ammonium phosphate, triple phosphate, and infection stones are synonyms), calcium phosphate (brushite), and calcium oxalate. Approximately half of calcium oxalate stones contain a minor component of calcium phosphate; the term "mixed calcium stone" is used when the calcium phosphate content exceeds 15% of the crystalline mass of the stone. Calcium oxalate and mixed calcium stones constitute about 80% of the stones encountered in the US, and the remaining four crystalline stone types are individually uncommon.

Many investigators stress the concepts of supersaturation chemistry in attempting to understand the pathogenesis of renal stone formation. These concepts are particularly useful when applied to the formation of the four "minor" stone types. For example, cystine is the least soluble of the naturally occurring amino acids, with a maximum solubility of about 300 mg/liter in aqueous solution. Patients with homozygous cystinuria commonly excrete 600–1400 mg of cystine per day so that their predisposition to cystine stone formation is self-evident. Further, there is a discrete causal relationship between cystinuria and the formation of cystine stones.

0066-4219/81/0401-0457$01.00

Although calcium oxalate is notoriously insoluble in aqueous solution, and the concepts of supersaturation chemistry may be applied to calcium oxalate stone formation, the pathogenesis of calcium stones is far more complex than the single and simple relationships between cystinuria and cystine stones. Thus the "risk factor" approach was introduced as a useful way of conceptualizing calcium stone formation in individual patients (2) or patient populations (3). The implied analogy to atherogenesis and coronary artery disease is fully intended, for in both instances statistical risk factors rather than single cause-and-effect relationships are recognized. Further, as in patients with coronary disease, multiple risk factors are often found in patients with recurrent stone disease.

The risk factors for calcium stone formation are (*a*) a family history of calcium stone formation, (*b*) dehydration, (*c*) certain medications (vitamins A, D, C, absorbable antacids, acetazolamide), (*d*) abnormal urinary acidification, (*e*) hyperoxaluria, (*f*) hyperuricosuria, and (*g*) hypercalciuria. In our experience, one or more of these risk factors is found in more than 90% of patients with recurrent calcium stone disease.

Hypercalciuria is an important risk factor for calcium stone formation and is identified in approximately 40% of patients with calcium stone disease (4). The upper panel of Figure 1 illustrates the rates of calcium excretion in normal subjects and in patients with calcium stone disease. Although there is a clear tendency toward higher rates of calcium excretion in the patients with stones, the normal and stone-forming populations are not discrete in terms of their rates of calcium excretion. This observation reflects the fact that many patients with calcium stones are not hypercalciuric; conversely, some 10% of the "normal" (defined simply by an absence of nephrolithiasis) population appear to excrete quantities of calcium in the "hypercalciuric" range. Nevertheless, increasing degrees of calciuria confer a definite statistical risk for calcium stone formation in a quantitative fashion (lower panel of Figure 1). The purpose of this review is to summarize current knowledge of renal calcium handling and then to review the pathophysiology and nomenclature of the various types of hypercalciuria.

RENAL CALCIUM TRANSPORT

Methods of Study

The renal handling of calcium can be studied at a variety of levels and by a variety of techniques including subcellular fractionation and transport techniques, the culture of isolated renal tubular cells, free flow micropuncture, isolated perfused nephron segments, isolated kidney or whole organism clearance studies, balance techniques, or combinations of the above. Each method had advantages and disadvantages. For example, the study of

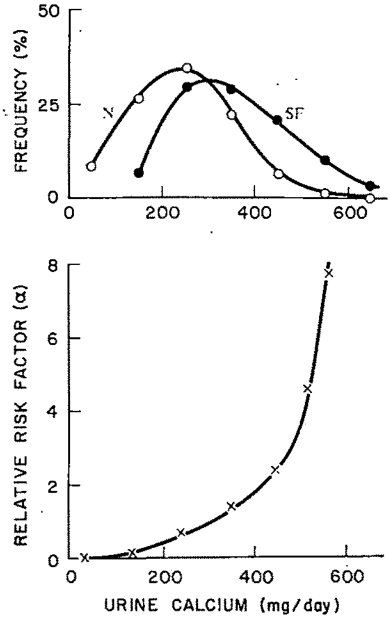


Figure 1 Urinary calcium excretion in normal (N) and calcium stone-forming (SF) individuals. Upper panel indicates the extensive overlap between these two populations and the rarity of extreme hypercalciuria. Lower panel indicates relative risk of calcium stone formation at any given level of calcium excretion. Adapted from Robertson et al (3).

calcium transport in subcellular organelles or components (isolated mitochondria, brush border membranes, basolateral membranes, etc) may introduce important alterations in function that are not operative in the intact cell. Similarly, micropuncture techniques are limited to certain superficial portions of the kidney; for example, proximal tubular fluid from deep, juxtamedullary nephrons cannot be sampled. In essence, currently available methods have not allowed the study of a number of relevant questions, and major information gaps therefore exist.

Calcium Transport At the Cellular Level

Although several important aspects of calcium transport through individual renal tubular cells remain to be elucidated, a model system can be described (Figure 2). Much of this information results from work in Borle's laboratory where cultured renal tubular cells were used (5).

The extracellular concentration of free calcium ions on both the luminal (brush border) and basolateral aspects of the cell approximates 10^{-3} M, while the cytosolic concentration of free calcium ions is approximately 10^{-7} M. The cytosol is electrically negative relative to the extracellular

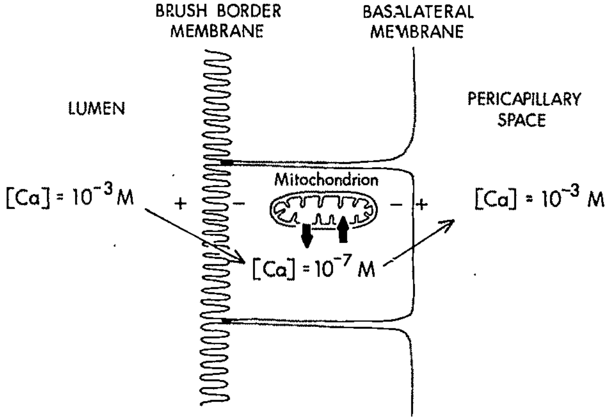


Figure 2 Calcium transport in a renal tubular cell. See text.

fluid. Calcium, therefore, enters the cell down a large electrochemical gradient and must exit against a similar gradient. Mitochondria display an enormous capacity for both releasing calcium into and removing calcium from the cytosol. Thus, calcium transport occurs across three different membranes—brush border, basolateral, and mitochondrial—and must be accomplished by membrane "pumps" in at least two of these three locations. The precise nature of these pumps is unknown.

Parathyroid hormone (PTH) stimulates net calcium transport across renal tubular cells. The mechanism whereby this is accomplished is uncertain and probably involves effects on mitochondrial calcium flux and possibly on brush border and/or basolateral membrane pumps as well. In some systems, the actions of PTH can be duplicated by analogues of cyclic AMP.

Transport of Calcium In the Nephron

Transport of calcium along the course of the nephron is depicted schematically in Figure 3. Direct sampling of glomerular filtrate in Bowman's capsule reveals calcium concentrations that are approximately 60% of those found in serum (5). This agrees well with the ultrafilterable calcium concentration as measured by dialysis techniques. Thus, slightly more than half of the calcium in serum is not bound to serum proteins and is filtered at the glomerulus.

Anatomically, the proximal tubule begins at Bowman's capsule and ends at the beginning of the thin descending limb of Henle's loop. It has both convoluted and straight (pars recta) portions. The distal tubule begins with the thick ascending limb of Henle's loop (diluting segment), continues through the distal convoluted tubule, and ends at the beginning of the collecting duct. Contemporary micropuncture terminology is governed by accessibility to the micropuncture pipette and may vary substantially from

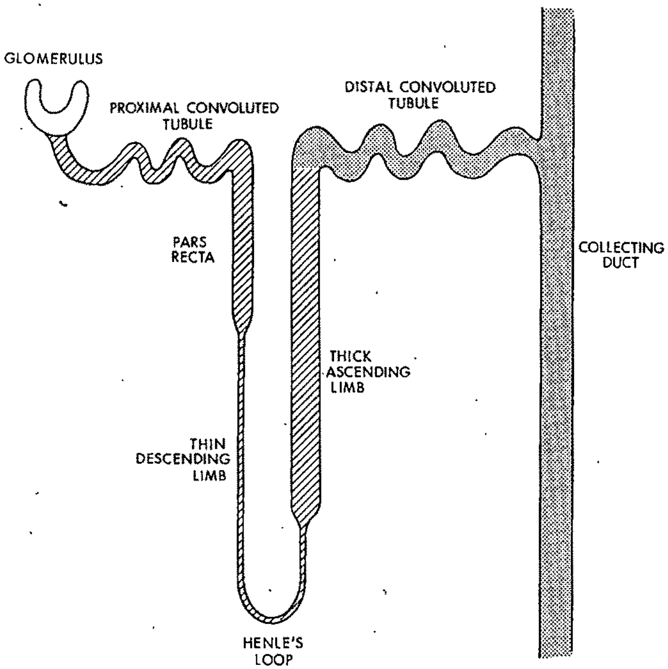


Figure 3 Diagramatic representation of a single nephron. Classical anatomic nomenclature is indicated on diagram. Current micropuncture nomenclature distinguishes proximal (*hatched*) from distal (*dotted*) nephron.

classical anatomic nomenclature. For example, "proximal" generally refers to the portion of the proximal tubule available to the micropuncturist and ends, therefore, at the beginning of the pars recta. "Distal" refers to all nephron segments distal to the earliest accessible sampling site in the distal convoluted tubule, and thus generally refers to distal convoluted tubule and the collecting duct. The remainder of nephron, i.e. the pars recta of the proximal tubule, and thin descending and ascending and the thick ascending portions of Henle's loop, may be considered "proximal" or may be further subdivided if the hairpin turn of Henle's loop is sampled.

Physiologically, calcium transport in the nephron is best divided into two basic regions (depicted by the hatched and dotted areas in Figure 3). Calcium transport occurring prior to the end of the thick ascending limb, i.e. before the distal tubule micropuncture site, is considered *proximal,* and that occurring after the distal sampling site is considered *distal.*

PROXIMAL NEPHRON Approximately 60% of the filtered load of calcium, sodium and water is reabsorbed prior to the proximal accessible site, and an additional 30% of calcium is reabsorbed by the beginning of the distal convoluted tubule. Thus, 90% of calcium reabsorption occurs proximally (5, 6). The predominant feature of proximal calcium reabsorption is its tight coupling to sodium reabsorption in these nephron segments. Factors that enhance sodium reabsorption in the proximal tubule (e.g. volume depletion) enhance calcium reabsorption as well. Factors that diminish proximal tubular sodium reabsorption (saline infusion, furosemide, acetazolamide) similarly reduce proximal calcium reabsorption (5, 6).

DISTAL NEPHRON By the end of the "proximal" nephron (i.e. by the beginning of the distal convoluted tubule), only 10% or approximately 1000 mg/day of the filtered load of calcium remains. Calcium reabsorption in this nephron segment, in contrast to proximal sites, is almost entirely dissociable from sodium handling (5, 6): factors that stimulate sodium reabsorption distally (e.g. aldosterone) have no influence on distal calcium handling; factors that enhance distal calcium reabsorption (e.g. PTH, chronic thiazide administration) have minor (if any) effects on sodium reabsorption. "Distal" calcium transport occurs against an electrochemical gradient and is thus active. Although direct evidence for a tubular maximum (Tm) or threshold for calcium reabsorption in the distal nephron is unavailable, in many ways the distal reabsorption mechanism functions physiologically as if it were Tm-limited. In this case, the Tm or "setpoint" for calcium reabsorption in the distal nephron is determined by the ambient circulating concentration of PTH. It is by this mechanism that distal calcium reabsorption is regulated to subserve systemic needs. Indeed, patients with chronic hypoparathyroidism represent a classic example of "renal" hypercalciuria even though the term seems oddly misapplied when used in this connection.

DIURETICS AND CALCIUM TRANSPORT Thiazide diuretics, used acutely, cause a natriuresis unassociated with changes in calcium excretion (6). Chronic thiazide use results in a fall in glomerular filtration rate (GFR) through sodium depletion and potentiates PTH effect on the distal tubule—both of which diminish net urinary calcium excretion (6). Loop diuretics (furosemide, ethacrynic acid) act primarily on the thick ascending limb of Henle's loop to block sodium chloride reabsorption, which diminishes proximal calcium reabsorption. The relatively greater increment in calciuresis than naturiuresis following furosemide suggests that the distal nephron may participate as well (6). Acetazolamide inhibits proximal sodium and calcium reabsorption but does not increase calcium excretion in the final urine. Metabolic acidosis resulting from chronic acetazolamide use may be associated with hypercalciuria. Potassium-sparing diuretics (spironolactone, triamterene) do not appear to influence calcium excretion.

CALCIUM HOMEOSTASIS IN THE WHOLE ORGANISM

Figure 4 is a schematic representation of systemic Ca homeostasis in a normal adult who is consuming a high-normal (1000 mg) calcium diet and who is in zero external calcium balance. The net quantity of calcium absorbed in the small intestine (175 mg/day) is equal to the total quantity absorbed (300 mg/day) less the amount lost via intestinal secretions (125 mg/day). The major mechanism of intestinal Ca absorption is active transport; only some 15% of Ca absorption proceeds by simple diffusion. In a nongrowing individual in zero calcium balance, the tightly coupled processes of bone resorption and bone mineralization are quantitatively equivalent and do not influence net calcium entry into the extracellular pool. The renal tubule plays the dominant quantitative role in normal minute-to-minute Ca homeostasis. The filtered load of calcium is approximately 10,000 mg/day. Of this quantity, some 98% is reabsorbed by the proximal and distal processes of calcium reabsorption. In essence, the control mechanisms are such that the normal net quantity of calcium absorbed via the intestine considerably exceeds systemic requirements for calcium. Thus, under normal circumstances urinary calcium excretion simply reflects this surplus of intestinally derived calcium.

The control mechanisms responsible for maintaining normal calcium homeostasis involve two hormones, parathyroid hormone and 1,25-$(OH)_2D$, acting on three target tissues: bone, kidney, and the small intestine. The mechanisms are sufficiently finely tuned that the ionized serum calcium normally varies by no more than ± 0.1 mg/dl throughout the day. The parathyroid cells are exquisitely sensitive to even minor fluctuations in serum calcium, so that the level of ionized calcium is the principal regulator of the rate of PTH secretion. The major target cells (and actions) of PTH are osteoclasts (osteoclastic bone resorption), the proximal renal tubule (reduction in phosphorus reabsorption), the renal cell containing 1-α-hydroxylase activity [increases 1,25-$(OH)_2D$ synthesis], and the distal nephron (increase in calcium reabsorption). PTH and hypophosphatemia represent two independent stimuli for 1,25-$(OH)_2D$ synthesis; in addition, the synthesis of 1,25-$(OH)_2D$ is strongly product inhibited. The principal target cells (and actions) of 1,25-$(OH)_2D$ are the small intestine (increase in active calcium and phosphorus absorption) and the osteoclast (increase in bone resorption). Minute-to-minute calcium homeostasis is maintained primarily by the effects of PTH on osteoclastic bone resorption and distal tubular calcium reabsorption. Increases in circulating 1,25-$(OH)_2D$ in response to a hypocalcemic challenge require some 24–48 hours to become maximal, so that the parathyroid–1,25-$(OH)_2D$ axis functions in a more

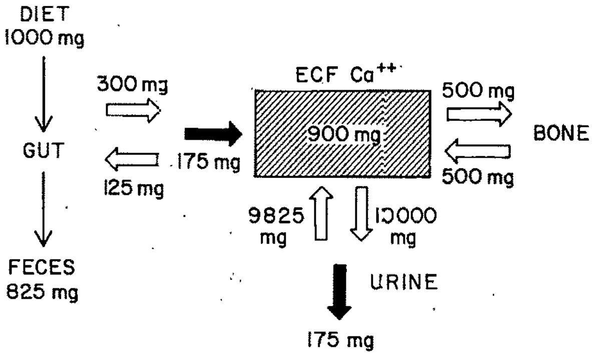


Figure 4 Total body calcium balance. See text. Net calcium movement is indicated by dark arrows.

subacute regulatory sense. Many investigators regard the function of this axis as analogous to the pituitary-adrenal axis, with the parathyroid glands serving as the primary endocrine gland, PTH as the trophic hormone, the kidney as the secondary endocrine gland, and 1,25-$(OH)_2D$ as the final, secondary endocrine product. It is the integrated effects of these various control mechanisms that are responsible for normal and abnormal systemic mineral homeostasis.

Figure 4 also indicates that relatively few fundamental mechanisms can lead to hypercalciuria: (*a*) a defect in renal tubular calcium reabsorption, (*b*) an increase in net intestinal calcium absorption, (*c*) an increase in bone resorption, or (*d*) some combination of these processes. Clearly, the two key variables determining the rate of calcium excretion are the integrated (i.e. fasting and postprandial) filtered load of calcium throughout the day and efficiency of renal tubular calcium reabsorption. Equally clearly, the quantity of calcium excreted in the urine must ultimately be derived from dietary calcium or bone mineral, or both.

HYPERCALCIURIA

Flocks, in 1939, initially noted the association between hypercalciuria and calcium stone formation (7). Albright et al, in 1953, coined the term idiopathic hypercalciuria (8), and the patient population so defined was reported in detail by Henneman et al in 1958 (9). The Henneman paper is a key historical reference in that it both defined the descriptive use of the term idiopathic hypercalciuria and broached the question of pathogenesis, thus initiating a controversy that still exists concerning the pathogenetic primacy and criteria for diagnosis of "renal" hypercalciuria, "absorptive hypercalciuria", and subtle or "normocalcemic" hyperparathyroidism.

Definition, Detection, and Incidence of Hypercalciuria

The seemingly simple issue of defining hypercalciuria and estimating its frequency in patients with calcium stone disease is, in fact, a matter of some consternation. The problems and variables involved are many and include (4) (*a*) the use of multiple arbitrary and inconsistent definitions of hypercalciuria in the literature, (*b*) reported differences in the population pattern of calcium excretion from one locale to another, (*c*) possible temporal and/or seasonal variability in the excretion of calcium by normal individuals and those with stone disease, (*d*) the influence of many dietary factors other than calcium on the rate of calcium excretion, and (*e*) the difficulty of assigning a statistically valid upper limit of calcium excretion in the normal population because of an apparent "positive skew" of values in this population (Figure 1). Of these various issues, dietary influences and the approach employed in screening for hypercalciuria are the most important.

Any quantitatively precise definition of hypercalciuria requires attention both to the various dietary factors influencing calcium absorption and excretion and to the dietary intake of calcium itself. Calcium absorption and/or excretion is increased by dietary protein, sodium, and carbohydrate and decreased by dietary oxalate and phosphorus. Of these factors, protein ingestion has the most striking quantitative effect on calcium excretion: increasing protein intake from 0.5 gm/kg/day (a restricted intake) to 2 gm/kg/day (a liberal intake) doubles calcium excretion in a normal individual (10). This effect appears to result from the metabolism of protein to fixed acid, with consequent effects on tubular calcium reabsorption. Sodium intake increases calcium excretion by approximately 25 mg per 100 meq increment in sodium excretion, so that wide swings in salt intake are required to significantly alter daily urinary calcium. The influences of carbohydrate, oxalate, and phosphorus appear to be quantitatively minor. The essential point is that any or all of these influences may affect the results of metabolic screening when the patient is on a "free" or undefined diet. As a practical point, this type of diet-induced "hypercalciuria" tends to be manifest only intermittently in a series of urine collections from a given patient, and almost always disappears when the patient is evaluated on a defined diet. Whether these dietary factors should be regarded as independent risk factors for hypercalciuria and stone formation is entirely unknown.

The second major dietary influence on calcium excretion is the quantity of calcium in the diet. In normal individuals, increasing calcium intake from low-normal (approximately 400 mg/day) to high-normal (approximately 1000 mg/day) has a modest impact on calcium excretion. In contrast, the majority of patients with hypercalciuria and stone disease display a marked increase in calcium excretion on a high-normal calcium intake, such that the segregation of these patients from normal tends to be far more clearcut when they are evaluated on an unrestricted rather than a restricted calcium intake (2).

The most useful single definition of hypercalciuria is a calcium excretion rate of >250 mg/day in females, >300 mg/day in males, or >4 mg/kg/day in either sex when the patient is screened on a defined 1000-mg calcium intake. These limits agree with those advocated by most investigators in the US and are virtually identical to those proposed by Lemann on the basis of an exhaustive literature search (10, 11).

Employing the dietary constraints and the upper normal limits described above, approximately 40% of patients with calcium stone disease will be found to have persistent hypercalciuria (1, 10, 11). As noted previously, the statistical and cause-and-effect relationships between calcium excretion and calcium stone formation are far less discrete than, for example, the relationship between cystine excretion and cystine stone formation. Nevertheless, both the presence and degree of hypercalciuria are statistically related to the risk of calcium stone formation. The relative risk is substantial at a calcium excretion rate of 300 mg/day and rises sharply at an excretion rate of ≥400 mg/day (Figure 1).

Classification of Hypercalciuria

Table 1 contains a classification of hypercalciuria. Of the secondary hypercalciurias, only the first six conditions or categories are clearly associated with an increased incidence of calcium stone disease. Medullary sponge kidney is listed as a "cause" of hypercalciuria, although the cause-and-effect relationship is unknown. Primary hyperparathyroidism is by far the most important of the secondary hypercalciurias, being identified on the average in 7.2% of 3084 patients with calcium stone disease reported in eight large series (12). The "subtle" variety of this disorder is easily overlooked and is given separate emphasis below. In the aggregate, the secondary hypercalciurias account for ~15–20% of patients presenting with calcium stone disease. "Dietary hypercalciuria" refers to two categories of hypercalciuria. The first is that associated with the consumption of large quantities of protein or other dietary factors that influence calcium excretion. Hypercalciuria is usually present only intermittently and its potential relationship to stone disease is unknown. "Calcium gluttony" can also produce hypercalciuria in an otherwise normal individual. The calcium intake must be massive, on the order of 3–4 gm daily, and the absorption is largely mediated by the nonsaturable or diffusional component of intestinal calcium absorption. Calcium gluttony is of limited importance as a cause of calcium stones and can be easily identified by dietary history.

Table 1 Classification of hypercalciuria

1. Secondary hypercalciuria
   a. Primary hyperparathyroidism
   b. Sarcoidosis
   c. Distal renal tubular acidosis
   d. Vitamin D excess
   e. Immobilization, rapidly progressive osteoporosis, malignant osteolysis
   f. Medullary sponge kidney (?)
   g. Glucocorticoid excess
   h. Thyrotoxicosis

1. Secondary hypercalciuria *(cont'd)*
   i. Acromegaly
   j. Furosemide administration

2. Dietary hypercalciuria

3. Idiopathic hypercalciuria
   a. Absorptive hypercalciuria
   b. Renal hypercalciuria
   c. Subtle primary hyperparathyroidism
   d. Resorptive hypercalciuria

Pathophysiological features of four categories or subtypes of idiopathic hypercalciuria are summarized in Table 2. This classification is a bit unconventional and is intended solely to emphasize the mechanisms of disease. In each instance, the apparent primary abnormality is indicated by an asterisk. It is of critical importance to recognize that all diagnostic classifications are based on operational definitions and assumptions and that pathogenetic mechanisms are of far greater importance than the semantics used to describe them.

The intestinal hyperabsorption of calcium in patients with absorptive hypercalciuria leads to postprandial increases in the extracellular concentration and filtered load of calcium, with attendant suppression of parathyroid function and spillage of the excessive load of calcium into the urine. The suppression of the distal tubular effects of PTH adds a "renal" component to the net hypercalciuria in this disorder. In addition, mounting evidence suggests that patients with absorptive hypercalciuria have an abnormality in bone mineralization and/or resorption and may thus have a "resorptive" component to their hypercalciuria as well (13). Patients operationally defined as having absorptive hypercalciuria appear to be pathogenetically heterogenous. There is evidence that at least three primary mechanisms may produce this basic syndrome. Shen et al (14) and Lemann et al (11) have proposed that many patients with absorptive hypercalciuria have a primary renal tubular phosphate leak, leading to a reduction in TmP/GFR and serum phosphorus concentrations and secondary increases in 1,25-$(OH)_2D$ synthesis and intestinal calcium absorption. Phosphorus repletion has been reported to partially correct these abnormalities. Other patients with absorptive hypercalciuria have increased circulating levels of 1,25-$(OH)_2D$ in the absence of reduced values of TmP/GFR or hypophosphatemia and are envisioned as having a primary abnormality in 1,25-$(OH)_2D$ production (15). Still another patient group displays intestinal hyperabsorption of calcium and clearly normal plasma concentrations of 1,25-$(OH)_2D$ and is regarded as having an enhanced intestinal sensitivity

Table 2 Pathophysiological features of four subtypes of idiopathic hypercalciuria[a]

	Urinary calcium excertion	GI calcium absorp-tion	Serium calcium	Renal phosphate threshold	1,25 Vitamin D	Para-thyroid function	Bone resorption
Absorptive hypercalciuria							
phosphate leak	+	+	N+	N–*	+	N–	N+
1,25-mediated	+	+	N+	N+	+*	N–	N+
intestinal	+	+*	N+	N	N	–	?N
Renal leak hypercalciuria	+*	+	N–	–	+	+	+
Subtle primary hyperpara-thyroidism	+	+	N+	–	+	+*	+
Resorptive hypercalciuria	+	–	N+	N	–	–	+*

[a] An asterisk indicates the apparent primary abnormality. N = normal; + = above normal; – = below normal.

to vitamin D or some equivalent disorder (16). The exact frequency of these various mechanisms in the population of patients with absorptive hypercalciuria is unknown.

Patients with renal (or "renal leak") hypercalciuria have a primary defect in renal tubular calcium reabsorption (16, 17). The site of this defect is unknown but is presumed to be in the distal nephron. The tubular losses of calcium lead to a state of secondary hyperparathyroidism, with resultant stimulation of 1,25-$(OH)_2D$ synthesis, increased calcium absorption, and reductions in TmP/GFR and serum phosphorus concentrations. Histomorphometric findings reveal a typical pattern of the "early" PTH response in bone, compatible with the state of relatively mild hyperparathyroidism. Thus, under steady-state conditions, patients with renal hypercalciuria have a combination of "renal," "absorptive," and "resorptive" components to their hypercalciuria. The pathophysiological abnormalities described are most clearly observed when patients with renal hypercalciuria are evaluated on a restricted calcium intake. Increasing the calcium intake to high-normal levels creates a new steady state with partial reversal of the secondary hyperparathyroidism and other secondary abnormalities; this reversal occurs at the expense of clearly accentuated hypercalciuria. Furosemide administration reproduces the pathophysiological pattern of renal hypercalciuria quite exactly (17).

The literature on the subject of "normocalcemic" primary hyperparathyroidism is justly regarded as controversial. The term itself is unfortunate, and "primary hyperparathyroidism with intermittent hypercalcemia" or "subtle" primary hyperparathyroidism are more descriptive and preferable terms. Mild hypercalcemia is the rule rather than the exception in patients with primary hyperparathyroidism and renal stone disease. The degree of hypercalcemia is often trivial, and a number of series have noted that the hypercalcemia may be present only intermittently in a large minority of these patients. Indeed, this intermittent pattern of hypercalcemia is a rather characteristic finding in patients with subtle hyperparathyroidism and is highly suggestive of the disorder. The disproportionality between the rather marked hypercalciuria and the mild or intermittent hypercalcemia in patients with subtle primary hyperparathyroidism has puzzled investigators for many years and has only recently been clarified by studies emphasizing the importance of circulating $1,25\text{-}(OH)_2D$ and intestinal calcium absorption in these patients (18).

In essence, patients with subtle primary hyperparathyroidism have a variant form of the disorder and represent the extreme end of the spectrum of the condition typified by mild hypercalcemia and disproportionate hypercalciuria. Although there is a clearcut resorptive component to the hypercalciuria in these patients, the $1,25\text{-}(OH)_2D$–mediated absorptive component is of far greater quantitative importance. In addition, the postprandial increases in serum calcium resulting from the intestinal hyperabsorption of calcium result in parathyroid suppression thus reversing the effects of PTH in the distal nephron. The net result is a pattern of calcium absorption and spillage of the absorbed load of calcium into the urine that is entirely analogous to the pattern of calcium absorption and excretion in patients with absorptive hypercalciuria.

Patients with resorptive hypercalciuria have a primary disorder associated with an increased rate of resorption of bone mineral. The mobilization of calcium into the extracellular pool suppresses parathyroid function, $1,25\text{-}(OH)_2D$ synthesis, and calcium absorption, and the filtered load of calcium is spilled into the urine. The best examples of resorptive hypercalciuria are distal renal tubular acidosis, immobilization, rapidly progressive osteoporosis, malignant osteolysis, and thyrotoxicosis. Resorptive hypercalciuria does not figure prominantly in the differential diagnosis of the vast majority of patients with idiopathic hypercalciuria. However, a resorptive component is present in the hypercalciuria of many patients with idiopathic hypercalciuria, as discussed above.

A thorough understanding of the pathophysiological features summarized in Table 2 serves two major purposes. First, it goes well beyond the simplistic semantics used to describe these syndromes and reveals that the hypercalciuria in each syndrome results from the interplay of a number of complex pathophysiological variables. Second, although there are many points of pathophysiological similarity among the various subtypes of idiopathic hypercalciuria, there are also distinctive points that must form the basis of differential diagnosis.

Diagnosis and Treatment

The risk factor approach to the differential diagnosis of patients with calcium stone disease has been described elsewhere and is not covered in this review (2). Similarly, a comprehensive summary of the diagnosis and treatment of the hypercalciurias is available elsewhere (1, 11). We will briefly outline our own diagnostic approach.

The presence and degree of hypercalciuria should be documented by two or more 24-hour urine collections performed on an unrestricted or high-normal (approximately 1000 mg) calcium diet known to contain normal quantities of protein (about 1 gm/kg/day) and sodium (about 100 meq/day). Employing an unrestricted calcium intake in screening for hypercalciuria is important because an increase in intestinal calcium absorption is common to each of the major variants of idiopathic hypercalciuria (excluding resorptive hypercalciuria).

The secondary hypercalciurias require a high index of suspicion, but diagnosis is usually straightforward on the basis of the history, examination, and routine clinical techniques. Complete distal renal tubular acidosis is a destructive but eminently treatable disorder and must be carefully ruled out. This is most easily achieved by equipping the patients with a roll of nitrazine paper and demonstrating his or her ability to achieve a urine pH of 5.3 or less on serial recordings. Treatment of the secondary hypercalciurias is disease-specific. Dietary hypercalciuria is identified by dietary history and by employing a defined diet in metabolic screening, as outlined above.

Having eliminated the secondary hypercalciurias and dietary hypercalciuria, the diagnosis of idiopathic hypercalciuria is achieved by exclusion. Although occasional patients with idiopathic hypercalciuria display an unusually prominant resorptive component to their hypercalciuria, most well-documented examples of resorptive hypercalciuria occur among the secondary hypercalciurias. Therefore, the principal goal of differential diagnosis is to segregate between the three major hypercalciuric subtypes illustrated in Table 2. No approach to this differential diagnosis has received universal endorsement. Indeed, criteria for diagnosis vary considerably among investigators, as do reported incidence figures for the frequency of the various hypercalciuric subtypes in the hypercalciuric population. However, there appears to be a growing consensus that patients with absorptive hypercalciuria constitute the clear majority of patients with idiopathic hypercalciuria.

Subtle primary hyperparathyroidism must be seriously considered in any patient with a mean serum calcium of $\geq$10.0 mg/dl on three determinations performed fasting and with careful attention to sampling and analytical technique. This diagnosis can be confirmed by further testing in approximately 80% of patients so defined. In our experience, the oral calcium tolerance test is an extremely useful diagnostic tool in these patients, and the combination of induced hypercalcemia and inappropriate suppression of parathyroid function makes specific diagnosis a relatively simple matter (19). Other investigators have recommended the simultaneous measurement of ionized serum calcium and immunoreactive PTH by a sensitive antiserum for diagnosis (20). The "thiazide challenge" test is discouraged on the grounds of nonspecificity. The treatment of subtle hyperparathyroidism is primarily surgical, and 80% of these patients harbor a small ($<$1 g) parathyroid adenoma.

Having excluded subtle primary hyperparathyroidism, the diagnosis rests between renal and absorptive hypercalciuria. Criteria for differential diagnosis include (*a*) coincident measurements of urinary calcium and parathyroid function in the fasting state, (*b*) coincident measurements of calcium and total cyclic AMP excretion in 24-hour samples [renal hypercalciuria is diagnosed by the combination of hypercalciuria and increased cyclic AMP excretion on a restricted calcium intake (21); absorptive hypercalciuria is diagnosed by the combination of hypercalciuria and suppressed cyclic AMP excretion in samples collected on a high-normal calcium intake], and (*c*) characteristic responses to a standard oral calcium challenge (2). Again, we have been impressed with the diagnostic accuracy of the oral calcium tolerance test and regard it as the most reliable diagnostic technique available.

The thiazide diuretics are clearly the drugs of choice in patients with renal hypercalciuria, and are capable of reversing partially or completely all the pathophysiological features of the syndrome (16, 17). There is no consensus regarding optimal medical therapy in patients with absorptive hypercalciuria. Oral phosphorus preparations and the thiazide diuretics are the two most widely used therapies in the US, but it is presently unclear whether either form of treatment can reverse the pathophysiological abnormalities in patients with absorptive hypercalciuria or its apparent variants. Patients are routinely instructed to force hydration and restrict calcium intake, but these measures have never been conclusively shown to be efficacious.

Many physicians regard either oral phosphorus or thiazides as their front-line therapy for idiopathic hypercalciuria and question the necessity for subclassifying hypercalciuria patients into renal and absorptive subpopulations. Pending (*a*) further information regarding natural history and the potential for clinical bone involvement in either of these patient groups, and (*b*) the results of detailed follow-up studies of the effects of therapy in well-characterized patient groups, this question must remain unanswered.

CONCLUSION

From the foregoing, it should be clear that many important questions regarding the pathogenesis of hypercalciuria remain unanswered. The area is further clouded by different diagnostic criteria, study conditions, and nomenclature employed by various investigators. Nonetheless, a clear conceptualization of the pathophysiology involved in the various hypercalciuric syndromes is an essential starting point in dealing with these patients, and will lead in most instances to a rational and successful therapeutic plan.

ACKNOWLEDGMENT

The authors wish to thank Dr. Fred Wright for his careful reading and helpful advice in preparing this manuscript.

Literature Cited

1. Coe, F. L. 1978. *Nephrolithiasis: Pathogenesis and Treatment,* p. 7. Chicago: Year Book Med.
2. Broadus, A. E., Thier, S. O. 1979. The metabolic basis of renal stone disease. *N. Engl. J. Med.* 300:839
3. Robertson, W. G., Peacock, M., Heyburn, P. J., Marshall, D. H., Clark, P. B. 1978. Risk factors in calcium stone disease of the urinary tract. *Br. J. Urol.* 50:499
4. Broadus, A. E. 1981. Nephrolithiasis. In *Endocrinology and Metabolism,* ed. P. Felig, J. Baxter, A. E. Broadus, L. A. Frohman. New York: McGraw-Hill. In press
5. Sutton, R. A. L., Dirks, J. H. 1978. Renal handling of calcium. *Fed. Proc.* 37:2112
6. Goldberg, M., Agus, Z., Goldfarb, S. 1976. Renal handling of phosphate, calcium, and magnesium. In *The Kidney,* ed. B. M. Brenner, F. C. Rector. Philadelphia: Saunders
7. Flocks, R. H. 1939. Calcium and phosphorus excretion in the urine of patients with renal or uretal calculi. *J. Am. Med. Assoc.* 113:1466
8. Albright, F., Henneman, P., Benedict, P. H., Forbes, A. P. 1953. Idiopathic hypercalciuria. *Proc. R. Soc. Med.* 46: 1077
9. Henneman, P. H., Benedict, P. H., Forbes, A. P., Dudley, H. R. 1958. Idiopathic hypercalciuria. *N. Engl. J. Med.* 259:802
10. Lemann, J., Adams, N. D., Gray, R. W. 1979. Urinary calcium excretion in human beings. *N. Engl. J. Med.* 301:535
11. Lemann, J. 1980. Idiopathic hypercalciuria. In *Nephrolithiasis (Contemporary Issues in Nephrology)*, ed. F. L. Coe, B. M. Brenner, J. H. Stein. New York: Churchill Livingstone
12. Broadus, A. E. 1980. Nephrolithiasis in primary hyperparathyroidism. In *Nephrolithiasis (Contemporary Issues in Nephrology), Vol. 5,* ed. F. L. Coe, B. M. Brenner, J. H. Stein. Edinburgh: Churchill Livingstone
13. Bordier, P., Ryckewart, A., Gueris, J., Rasmussen, H. 1977. On the pathogenesis of so-called idiopathic hypercalciuria. *Am. J. Med.* 63:398
14. Shen, Fri., Ivey, J. L., Sherrant, D. J., Neilsen, R. L., Haussler, M. R., Baylink, D. J. 1978. Further evidence supporting the phosphate leak hypothesis of idiopathic hypercalciuria. *Adv. Exp. Med. Biol.* 103:217
15. Kaplan, R. A., Haussler, M. R., Deftos, L. J., Bone, H., Pak, C. Y. C. 1977. The role of $1\alpha 25$ dihydroxyvitamin D in the mediation of intestinal hyperabsorption of calcium in primary hyperparathyroidism and absorptive hypercalciuria. *J. Clin. Invest.* 59:756
16. Pak, C. Y. C. 1979. Physiological basis for absorptive and renal hypercalciuria. *Am. J. Physiol.* 237:F415
17. Coe, F. L., Canterbury, J. M., Firpo, J. J., Reiss, E. 1973. Evidence for se-

condary hyperparathyroidism in idiopathic hypercalciuria. *J. Clin. Invest.* 52:134

18. Broadus, A. E., Horst, R. L., Lang, R., Littledike, E. T., Rasmussem, H. 1980. The importance of circulating 1,25 dihydroxyvitamin D in the pathogenesis of hypercalciuria and renal stone formation in primary hyperparathyroidism. *N. Engl. J. Med.* 302:421
19. Broadus, A. E., Horst, R. L. Mahaffey, J. E., Rasmussen, H. 1980. Primary hyperparathyroidism with intermittent hypercalcemia: serial observations and simple diagnosis by means of an oral calcium tolerance test. *Clin. Endocrinol.* 12:225
20. Muldowney, F. P., Freaney, R., McMullen, J. P., Towers, R. P., Spillane, A., O'Connor, P., O'Donohoe, P., Maloney, M. 1976. Serum ionized calcium and parathyroid hormone in renal stone disease. *Q. J. Med.* 45:75
21. Broadus, A. E., Dominguez, M., Bartter, F. C. 1978. Pathophysiological studies in idiopathic hypercalciuria: use of an oral calcium tolerance test to characterize distinctive hypercalciuric subgroups. *J. Clin. Endocrinol. Metab.* 47:751

Ann. Rev. Med. 1981. 32:475–90

THE NATURAL HISTORY AND TREATMENT OF OVARIAN CANCER[1]

❖7413

Dan. L. Longo, M.D. and Robert C. Young, M.D.

Medicine Branch, Division of Cancer Treatment, National Cancer Institute,
Bethesda, Maryland 20205

INTRODUCTION

Every fifty minutes someone in the United States dies from ovarian cancer. There are 17,000 new cases every year, about one fifth of all cases of ovarian masses. Malignant tumors of the ovary may arise from epithelial, stromal, or germinal structures. Since over 85% of all ovarian cancers are of epithelial origin, they are the major focus of this discussion and are termed "ovarian cancer." The reader is referred to other recent reviews for a discussion of ovarian tumors of stromal or germinal cell origin (1–3).

The etiology of ovarian cancer is unknown; however, epidemiologic studies implicate hormonal and environmental influences (4). Nulliparous women and those with a history of difficulty conceiving children have a higher incidence than multiparous women (5). Women with ovarian cancer have three to four times greater risk of developing breast cancer (6), an association that could suggest common etiologic factors. The evidence implicating environmental influences is equally circumstantial. Ovarian cancer, virtually unheard of prior to the Industrial Revolution, has its greatest incidence in industrialized countries. Migrants to high incidence areas and their subsequent offspring develop ovarian cancer with a frequency approaching that of natives of high incidence areas (7). Environmental exposure to asbestos and talc may play an etiologic role, although firm proof is lacking (8). Female asbestos workers have a much higher than expected incidence of ovarian cancer (9), and one group of investigators has detected talc particles imbedded in human ovarian cancer specimens (10). There are rare instances of families with many ovarian cancer victims, which might support a role for genetic influences in the development of ovarian cancer (11). However, no conclusive data on etiology exist.

---

[1] The US Government has the right to retain a nonexclusive royalty-free license in and to any copyright covering this paper.

In spite of our poor understanding of the factors responsible for the development of ovarian cancers, a number of careful studies have identified prognostic factors and important aspects of the natural history of the disease, and defined necessary staging procedures. This new understanding forms a foundation for therapeutic strategies designed to control ovarian cancer.

NATURAL HISTORY

Ovarian cancer is generally a disease of postmenopausal or perimenopausal women that remains predominantly intraperitoneal throughout its course. Unfortunately, it tends to remain clinically silent until the disease is advanced. In one large series of patients with ovarian cancer (12), initial signs and symptoms included abdominal pain (56%), abdominal swelling (46%), 10-lb weight change (usually loss) (31%), and abnormal vaginal bleeding (22%). The majority of these signs and symptoms already indicate disease spread out of the pelvis. Ovarian cancer kills its victims by causing progressive inanition and gastrointestinal and urinary tract obstruction that cannot be surgically reversed because of extensive intraabdominal carcinomatosis.

Ovarian cancer spreads contiguously within the peritoneal cavity, through the flow of tumor-cell-containing ascitic fluid, and through the rich lymphatic drainage of the ovary. The lymphatics of the ovary drain primarily through the paraaortic chain and up to the diaphragm with secondary drainage to iliac nodes. The presence of ovarian cancer cells in the nodes alters the flow of lymph and produces hydrostatic pressure changes resulting in ascites formation (13). This ascitic fluid bearing tumor cells bathes the peritoneal cavity and hemidiaphragms resulting in plaque-like tumor growth on virtually any peritoneal surface. Distant lymphatic and hematogenous metastases are seen occasionally.

PROGNOSTIC FACTORS AND STAGING

The FIGO staging scheme for ovarian cancer is shown in Table 1. The five-year survival of patients treated with surgery alone with Stage I disease (limited to ovaries) is about 67%; for Stage II disease (limited to true pelvis) about 25%; and for Stage III (intraperitoneal disease) and Stage IV disease (spread to liver parenchyma or extraperitoneal sites) about 5% (14). Thus, the stage of disease is a key prognostic factor.

Although accurate staging hinges on a careful surgical procedure, a variety of noninvasive and minimally invasive tests are useful in defining the

Table 1 Staging scheme of the International Federation of Gynecology and Obstetrics for ovarian cancer and survival rates following surgery ± irradiation and/or single agent chemotherapy

Stage	Five-year survival (%)
I. Growth limited to ovaries	60
A. Growth limited to one ovary; no ascites	
B. Growth limited to both ovaries; no ascites	
C. Tumor Stage IA or IB plus ascites or malignant cells in peritoneal washings	
II. Growth involving one or both ovaries with pelvic extension	39
A. Extension or metastases or both to uterus or tubes or both	
B. Extension to other pelvic tissues	
C. Tumor Stage IIA or IIB plus ascites or malignant cells in peritoneal washings	
III. Growth involving one or both ovaries with intraperitoneal metastases outside the pelvis or positive retroperitoneal nodes, or both; tumor limited to true pelvis with histologically proven malignant extension to small bowel or omentum	6
IV. Growth involving one or both ovaries with distant metastases; pleural effusion must contain malignant cells to indicate Stage IV disease; parenchymal liver metastases indicate Stage IV disease	4

extent of disease. Bipedal lymphangiography can detect clinically unrecognized paraaortic node involvement. About 25% of patients thought at surgery to have Stage I or II disease can be shown by lymphangiography to have extrapelvic tumor involving the paraaortic nodes (15). Intravenous pyelography and barium enema examinations may aid in the detection of ureteral, bladder, and bowel encroachment. Ultrasonography may be useful in detecting small amounts of ascites, nodal enlargement, or masses not felt on physical examination. Abdominal computerized tomography is being studied as a noninvasive replacement for more invasive procedures.

Peritoneoscopy allows inspection of peritoneal surfaces including the diaphragms, biopsy of suspicious lesions, and sampling of fluid for cytological evaluation. About 25% of patients with presumed localized disease undergoing peritoneoscopy were upstaged when subdiaphragmatic nodules undetected at surgery were found (16). Examination of peritoneal washing cytology in patients with no gross abnormality changed the stage of 12% of patients in another series (17). It is clear that peritoneoscopy is a useful adjunct to surgery in defining the extent of disease and thus in determining one important prognostic factor.

A second prognostic factor in ovarian cancer is the amount of residual disease after debulking surgery. Aggressive cytoreductive surgery, which reduces residual tumor to deposits less than 1.5 cm in diameter, results in improved survival (18). In fact, complete resection of tumor in patients with Stage III disease results in longer survival than incomplete resection of Stage II disease. While we have described extent of residual disease as a separate prognostic factor, stage and extent of residual disease are clearly linked. Historically, staging has been one approach to defining the amount of residual disease. The data on debulking surgery simply reinforces the importance of assessing the patient's tumor burden, and points out that we must not adhere too rigidly to artificially derived staging schemes proposed prior to the development of effective staging and therapeutic measures.

A third major prognostic factor in ovarian cancer is the histologic grade of the tumor. Several studies (19, 20) employing either the pattern of tumor growth or Broder's classification for the degree of cellular atypia within the tumor have recorded significant differences in survival between well- and poorly differentiated tumors within each stage of disease. In one study (19), overall five-year survival of Stage II patients was 38%, but when patients were stratified according to histologic grade, those with little atypia (grade 1) had an 80% five-year survival while those with highly anaplastic tumors (grades 3 and 4) had only a 10% five-year survival.

THE ROLE OF SURGERY IN OVARIAN CANCER

Despite all that has been learned about the natural history and patterns of spread in ovarian cancer, the majority of patients referred to Roswell Park Memorial Institute (83%) (21) and the National Cancer Institute (75%) (22) have had subumbilical or other limited incisions through which adequate examination of the upper abdomen would be virtually impossible. An adequate staging operation, in addition to total hysterectomy and bilateral salpingo-oophorectomy, should include (*a*) resection of as much disease as technically feasible to minimize residual disease and assure adequate tissue specimens for histologic grading; (*b*) examination and biopsy of paraaortic nodes; and (*c*) careful inspection of all peritoneal surfaces including the posterior gutters, both hemidiaphragms, and the liver surface.

The present role of surgery in ovarian cancer is primarily to determine the pathological stage, reduce tumor bulk, and assess the completeness of the response to other treatment modalities. The one setting in which surgical resection alone may be adequate treatment is in Stage IA well-differentiated (low grade) ovarian cancer in which 87% of patients survive five years without additional treatment (19).

RESTAGING

While accurate staging at initial surgery is critical, periodic restaging is also important both to assure that the 25% of patients who present with Stage

I or II disease are not progressing, and to evaluate the effects of treatment in the 75% of patients who present with advanced disease. The same features of the disease that cause the majority of patients to seek medical attention only after the disease has become advanced, often make detection of residual disease by clinical examination alone impossible. Certainly data accumulated to date have stressed the differences between clinical and pathologically documented complete responses. Between one third and one half of patients thought to be disease-free on clinical grounds are actually found to be pathologically disease-free at surgery (23). Those that were disease-free had a 72% five-year survival after restaging while those with tumors larger than 2 cm at "second look" operation had only a 17% five-year survival (23).

The need for a restaging evaluation proceeding all the way to repeat laparotomy may be significantly reduced by the judicious use of the same noninvasive and minimally invasive tests employed in the initial staging of the patient. Lymph nodes with persistently abnormal lymphangiographic contours may be safely sampled nonsurgically by percutaneous aspiration of the lymph node under fluoroscopic guidance (24). Peritoneoscopy may eliminate the need for laparotomy in over one third of patients (22, 25). We have found 24 of 66 patients (36%) in clinical complete remission after chemotherapy had findings of residual disease at peritoneoscopy necessitating continuation of therapy (22). Another group found 8 of 22 patients (36%) to have documented evidence of persistent disease at peritoneoscopy (25). The absence of findings of residual disease at peritoneoscopy cannot be a substitute for definitive restaging laparotomy, however. In one series, 2 of 10 patients (20%) with negative peritoneoscopy had tumor at laparotomy (25), and in another series (22), 58% of patients with negative findings at peritoneoscopy were found to have residual disease at operation. Future trials stratified by prognostic factors and incorporating the principles of accurate staging and restaging should provide more meaningful data on therapeutic efficacy.

TREATMENT OF OVARIAN CANCER

It is difficult to assess the value of a particular treatment regimen for ovarian cancer because the vast majority of studies have failed (*a*) to assure a balance of patients with similar prognostic factors between treatment arms, (*b*) to adequately stage patients prior to treatment, (*c*) to document pathologic disease-free status at the end of treatment, and (*d*) to carefully follow patients to detect recurrent disease at the earliest possible time. As a result, the clinician is unable to cull from the published literature the optimum therapy for patients with this disease. We shall attempt to summarize the most important studies done to date and highlight what useful information has been learned.

Radiation Therapy

Trials of radiation therapy in both early and advanced stage disease have been reported. Techniques of administration of radiation therapy, though beyond the scope of this review, vary widely in their efficacy and incidence of complications. It is clear, however, that the best results with radiation therapy have been obtained with specialized approaches perfected and applied at only a small number of referral centers.

Several studies have attempted to document that postoperative pelvic radiation improves survival of Stage I patients, but none has succeeded (26–28). In fact, one randomized study showed twice as many relapses (35%) in the treated as in the untreated (17%) patients (28).

None of the prospective stratified randomized trials document a role for pelvic radiation in Stage II ovarian cancer. In one study (26), Stage II patients given pelvic radiation had a 39% five-year survival as compared to 24% in the untreated group. This minimal difference is certainly not reassuring that radiation is helpful in presumed localized disease.

In view of the patterns of spread of ovarian cancer, Bush and his colleagues at the Princess Margaret Hospital in Toronto use a unique total abdominal radiotherapy technique to deliver effective doses of radiation (2200 rads) to the upper abdomen and diaphragms without untoward renal, hepatic, and bone marrow toxicity. They compared pelvic radiation (4500 rads) alone to pelvic radiation plus fixed low dose daily chlorambucil, or the unique total abdominal radiation approach in patients with Stages I and II disease and in those Stage III patients with minimal residual disease (29). The patients receiving pelvic radiation alone had 50% four-year survival. The addition of chlorambucil resulted in 55% four-year survival, and the addition of abdominal radiation resulted in 81% four-year survival. Unfortunately, the patients in this study were not adequately staged and stratified for extent of residual disease and the chemotherapy was not optimally administered. However, the study does suggest that pelvic radiotherapy results can be improved upon by expanding the ports to include areas at high risk from spreading disease. Interestingly, the Stage III patients included in this trial were those with the best chance of responding to combination chemotherapy (30, 31) (see below).

At Stanford, another selected group of Stage III patients with paraaortic node involvement alone were treated with 4500 rads to the nodes in addition to abdominal and pelvic ports (15). Patients treated in this fashion had a 53% five-year survival, significantly better than historical controls. Whether or not inclusion of the medial two thirds of the hemidiaphragms in the radiation ports will improve the results awaits trial (32).

The role of radiotherapy in the treatment of ovarian cancer remains to be established. No study has yet defined a subgroup of patients for which radiation therapy is uniquely the best treatment.

Chemotherapy

There are many anticancer drugs with demonstrable activity as single agents in the treatment of ovarian cancer (Table 2). The assembly of such drugs into active combinations represents a significant improvement in the treatment of victims of this disease. Today the majority of patients with

Table 2 Single agents active in advanced ovarian adenocarcinoma

Drug	Schedule	Patients (no.)	Response rate (%)
Alkylating agents			
Melphalan	0.2 mg/kg/day × 5 p.o. or i.v. q 35 wk	494	47 (20 complete)
Chlorambucil	0.2 mg/kg/day p.o.	280	50
Thiotepa	10 mg/day × 15 courses i.v.	144	65
Cyclophosphamide	50–150 mg/day p.o.	126	49
	400 mg/day × 4 days i.v. then 50–150 mg/day p.o.	104	37
Mechlorethamine	0.2 mg/kg/day × 2 i.v. then chlorambucil 8–14 mg/day p.o.	81	35
Antimetabolites			
5-Fluorouracil	15 mg/kg/day × 5 then 7.5 mg/kg q.o.d./34 wk	81	32 (18–20)
	15 mg/kg i.v./wk	21	33
Methotrexate	5 mg/day × 5–10 p.o. or i.v. 3–4 wk	16	25
	1–7.5 gm/m^2 i.v. wk with leukovorin rescue	23	13
Vinca alkaloids			
Vinblastine	0.1–0.15 mg/kg/day × 1–3 i.v.	16	13
Miscellaneous			
Hexamethylmelamine	8 mg/kg/day p.o. or 6 mg/kg/day p.o. × 14 days q 4 wk	53	41
Doxorubicin hydrochloride	30 mg/m^2/day × 3 i.v.	18	28
	50 mg/m^2 i.v. q 3 wk	33	36
Progestogens	200–600 mg/wk i.m.	50	10
Cisplatin	30 mg/m^2 daily i.v. × 3 q 28 days	34	27

advanced ovarian cancer can count on surviving longer as a result of chemotherapy, making it the treatment of choice in advanced disease. The role of chemotherapy in early stage disease is still controversial. Let us examine the data on the effectiveness of chemotherapy in each stage of disease.

STAGES I AND II (EARLY STAGE) The role of chemotherapy in the adjuvant setting is unsettled. Several randomized trials have been undertaken, but long-term follow-up is inadequate at this point to firmly establish the most appropriate management for the one patient in four presenting with disease limited to the pelvis.

Adjuvant chemotherapy appears to be at least as good as intracavitary radioisotopes (33) at limiting recurrent disease. One single-arm study employing adjuvant chlorambucil reported 94% five-year survival in 14 patients with Stage I disease (34). Comparisons of radiation therapy and single alkylator agent treatment show chemotherapy to be as good or better. In a study randomizing Stage I patients to melphalan or pelvic plus abdominal radiation, no significant survival differences have been noted in six years of follow-up (28). Eighty-five percent of irradiated patients are disease-free as compared to 90% of melphalan-treated patients (28).

The Ovarian Cancer Study Group, in conjunction with the Gynecologic Oncology Group, has two ongoing studies that should help decide the most appropriate management for early stage disease. They are comparing melphalan to observation in Stage I patients with well- or moderately differentiated grades of tumor, and are comparing melphalan to intracavitary ^{32}P in Stage I patients with ascites or tumor outside the ovarian capsule or with poorly differentiated grades of tumor. There is no survival data available as yet from these studies but they should help define the utility of adjuvant therapy in Stage I disease.

With careful staging, the number of patients who remain with Stage II disease is quite small (16). Thus, in most studies in which careful staging has been done, Stage II and III patients have similar survivals. Two recent studies have examined the effects of single-agent chemotherapy and radiotherapy in Stage II disease. In one study of patients not rigorously staged, treatment consisted of pelvic radiation, pelvic radiation followed by chlorambucil, or pelvic plus total abdominal radiotherapy with fields extended to include the diaphragms (35). There were no survival differences among the groups. The group at the M. D. Anderson Hospital randomized carefully staged patients with minimal residual disease to receive either melphalan or pelvic plus abdominal radiotherapy (28). With 3–6 year follow-up, the percentages of disease-free patients are 58% and 55%, respectively. Thus, in Stage II disease, clearcut treatment recommendations are not yet possible.

The Ovarian Cancer Study Group, in conjunction with the Gynecologic Oncology Group, is comparing pelvic radiotherapy plus melphalan to melphalan alone in patients with Stage II disease. However, patient accrual to this protocol has been slow and definitive answers are probably ten years away. It is clear that true Stage II patients need some form of adjuvant treatment but no treatment modality is clearly superior at present. It is distressing that half these patients die of their disease within five years of diagnosis.

STAGES III AND IV (ADVANCED STAGE) In advanced stage disease, the prolongation of survival by melphalan employed as a single agent has established chemotherapy as the treatment of choice: 35–65% of patients respond to melphalan and 20% of the responding patients will survive five years (36). The history of chemotherapy has taught us that when several independently active drugs are available, using them in combination often improves survival (37).

Several studies have employed combination chemotherapy in previously untreated patients (Table 3). One study reported a 60% (25 of 41 patients) response rate from a regimen employing actinomycin D, 5-fluorouracil, cyclophosphamide, vincristine, and prednisolone. Median duration of response was eight months with only two patients disease-free for more than one year (38). A four-drug combination regimen consisting of cyclophosphamide, hexamethylmelamine, doxorubicin hydrochloride, and *cis*-platinum gave a 67% (8 of 12 patients) response rate, but the short duration of follow-up does not allow assessment of survival prolongation (39). These and other studies of combination regimens that have been carried out do not take into account histologic grade or extent of residual disease, and none of these studies compares combination to single-drug regimens. They do not address the key question of whether combination regimens are superior to single agents in the treatment of advanced disease.

Two studies comparing actinomycin D, 5-fluorouracil, and cyclophosphamide (ACFUCY) to single-agent L-PAM failed to translate slightly higher response rates to the combination therapy into prolonged survival (40, 41). The doxorubicin hydrochloride plus cyclophosphamide regimen has been compared to single-agent cyclophosphamide (31) and single-agent L-PAM (42) in separate studies. In neither study was the combination better than the single agent in patients with bulky residual disease, but in the former study doxorubicin hydrochloride plus cyclophosphamide was significantly more effective than single-agent cyclophosphamide in patients with minimal residual disease (31).

The first prospective randomized trial to demonstrate the superiority of a combination regimen over standard single-agent alkylator therapy was

Table 3 Combination chemotherapy for ovarian cancer

Regimen	Responses	Complete responses	Duration of response/survival	Reference
Actinomycin D	60%	5%	8 months/—	38
5-FU	25/41	2/41		
Cyclophosphamide				
Vincristine				
Prednisolone				
Cyclophosphamide	67%	25%	N/A	39
Hexamethylmelamine	8/12	3/12		
Doxorubicin hydrochloride				
Cis-platinum				
Doxorubicin hydrochloride	36%	6%	6 months/12 months	31
Cyclophosphamide	13/36	2/36		
vs				
Cyclophosphamide	31%	0	6 months/12 months	
	11/35			
Doxorubicin hydrochloride	73%	54%	—/10 months	42
Cyclophosphamide	8/11	6/11		
vs				
L-PAM	29%	7%	—/10 months	
	4/14	1/14		
Hexamethylmelamine	75%	33%	>30 months/29 months	30
Cyclophosphamide	30/40	13/40	complete remission	
Methotrexate				
5-FU				
vs				
L-PAM	54%	16%	>25 months/17 months	
	20/37	6/37	complete remission	
Actinomycin D	53%	28.6%	—/>9 months	40
5-FU	26/49	14/49		
Cyclophosphamide				
vs				
L-PAM	35%	18.4%	—/>9 months	
	17/49	9/49		

done at the National Cancer Institute and compared hexamethylmelamine, cyclophosphamide, methotrexate, and 5-fluorouracil (Hexa-CAF) to L-PAM (30). Hexa-CAF-treated patients had a higher response rate (75% vs 54%), more complete remissions (33% vs 16%), and a prolonged median survival (29 vs 17 months). Patients with minimal residual disease after initial surgery had a higher overall response rate than patients with bulky disease (84% vs 53%) and histologic grade was confirmed to be an important factor in predicting response (20). The effectiveness of the Hexa-CAF regimen has subsequently been confirmed by other investigators (43).

More recently, preliminary reports claim complete response rates of up to 52% (44) for regimens employing *cis*-platinum. However, follow-up is too short at this time to determine whether these complete responses will have an impact on survival.

RESPONSE TO CHEMOTHERAPY AFTER INITIAL TREATMENT FAILURE The optimistic response rates reported for platinum-containing combinations acquire greater importance when it is recognized that initial treatment failures are remarkably resistant to second-line chemotherapy. For example, Hexa-CAF, which produces a 75% response rate in previously untreated patients, caused only partial responses in 23% (3 of 13 patients) after relapse from single-agent chemotherapy or radiotherapy (43). In 360 patients at M. D. Anderson Hospital who received a variety of treatments (including combinations) after failure of primary therapy, only 6% responded to second- or third-line drugs (45). Doxorubicin hydrochloride, which has a 36% response rate in untreated ovarian cancer patients, produced short-term partial responses in 3 of 56 patients who had failed initial chemotherapy (46, 47). A variety of combinations have been tried as salvage regimens for primary treatment failures. Hexamethylmelamine plus 5-fluorouracil (36%) (48), doxorubicin hydrochloride plus *cis*-platinum (42%) (49), cyclophosphamide (or 5-fluorouracil) plus doxorubicin hydrochloride, hexamethylmelamine, and *cis*-platinum (48–67%) (50–52) all have moderate response rates. However, complete responses are extremely rare, duration of follow-up is short, and the toxicity of the combination regimens is substantial in these previously heavily treated patients. Thus, until less toxic and more effective salvage regimens are found, complete response to initial therapy will be the major determinant of survival.

New Approaches to the Treatment of Ovarian Cancer

The use of nonspecific immunostimulants to enhance the effects of chemotherapy in ovarian cancer has been proposed. Although most of the studies on this point suffer from being single-arm trials in heterogenous patient populations, two studies suggest that immunotherapy might be useful. The Gynecologic Oncology Group compared melphalan plus intravenous *Corynebacterium parvum* to a historical group who received melphalan alone (53). More patients receiving immunotherapy responded (53% vs 29%) and the progression-free interval (12 vs 6 months) and survival (24 vs 12 months) were twice as long as patients receiving melphalan alone (53). The Southwest Oncology Group randomized untreated patients to doxorubicin hydrochloride plus cyclophosphamide with or without bacillus Calmette-Guerin (BCG) given by scarification. The patients receiving BCG had a higher response rate (53% vs 36%), more complete responses (12% vs 2%), and a median survival nearly twice as long (24 vs 13 months) as patients receiving chemotherapy alone (54). Complete information on the comparability of prognostic factors in each treatment group is not available for either of these trials. However, such studies provide hope that immunoadjuvants might improve response rates and survival in advanced disease when added to effective combination chemotherapy, and raise the question of whether immune therapy might be useful in early disease.

The development of the Tenckhoff catheter, a chronic indwelling intraperitoneal catheter, has allowed reconsideration of administering active agents intraperitoneally. Because ovarian cancer generally remains confined to the peritoneal cavity, intraperitoneal administration should allow higher concentrations of drugs to be maintained in direct contact with the tumor for longer periods and with less systemic toxicity than intravenous administration (55). Phase I trials with intraperitoneal methotrexate have shown that concentration gradients of 10–30-fold between peritoneum and plasma drug levels can be achieved (56). Though toxicities such as chemical peritonitis and myelosuppression have been noted, intracavitary administration of active drugs such as 5-fluorouracil (57) and doxorubicin hydrochloride (58) and drug combinations may be a unique approach to the treatment of ovarian cancer, especially in the adjuvant setting or in patients with minimal residual disease.

The capacity to establish in vitro conditions for the proliferation of human tumor stem cells (59) has allowed the development of an in vitro prediction of clinical responses to particular agents. The ability of a drug to inhibit colony formation in soft agar after a brief exposure has been found in 44 ovarian cancer patients to predict in vivo resistance with 99% accuracy and in vivo sensitivity with 64% accuracy (60). The earliest application of this technique will be to individualize salvage therapy in patients relapsing during primary therapy on the basis of in vitro susceptibility testing, much the same way that antibiotic susceptibility testing of bacterial isolates guides clinical decision-making in infectious diseases.

New drug development and the study of useful drugs that have not been adequately tested in ovarian cancer are other areas in which research may bring important advances in treatment. Bleomycin, procarbazine, and streptozotocin, whose dose-limiting toxicities are not myelosuppression, are candidates for more combinations. Mitomycin C and its derivative, porfiromycin, have not been carefully studied, and vinblastine deserves restudy in light of its activity as a salvage agent, as has been shown using the stem cell assay (60). In addition, the radiation- and chemotherapy-sensitizer misonidazole or a congener may be useful in the next generation of ovarian cancer treatment regimens.

Basic science investigation has recently uncovered an interesting finding that may ultimately be of clinical utility. Ovarian carcinomas arise in cells derived embryologically from Mullerian epithelium. In males, who have an extraordinarily low incidence of epithelial gonadal tumors, Mullerian regression factor (MRF) causes involution of Mullerian epithelium during sexual differentiation. MRF purified from male embryos has cytotoxic effects on human ovarian cancer cells in vitro (61). If MRF becomes available in bulk quantity (perhaps through recombinant DNA technology), it would be of interest to test its efficacy in patients with ovarian cancer.

SUMMARY

Several advances in the understanding of the biology and treatment of ovarian cancer over the past five years have changed the pessimistic view of this disease that was formerly commonplace.

A cornerstone of the management of patients with ovarian cancer is careful pathological staging both prior to and at the cessation of therapy. Patients with all grades of Stage I lesions except the well-differentiated ones should have some form of adjuvant therapy, although the best choice is still unclear. Similarly, Stage II patients, though few in number, have only a 50% chance of surviving five years without some adjuvant treatment. Since the best choice is unclear, it would be ideal if all such patients could be treated in a study setting. Every early stage patient treated off study slows the identification of definitive therapy.

For the majority of patients with advanced disease, the treatment of choice is aggressive debulking surgery followed by combination chemotherapy. After six cycles of therapy, patients should be restaged sequentially moving all the way to repeat laparotomy, if necessary, to find residual disease before therapy is discontinued. If no disease is found, we randomize patients to either careful follow-up or six cycles of intraperitoneal adjuvant therapy in an effort to prevent relapse.

If residual disease is found at restaging (but less disease than at the beginning of treatment), the patient is given six more cycles of combination chemotherapy, though often at reduced doses because of cumulative bone marrow toxicity. After the twelfth cycle of chemotherapy, restaging is repeated. If disease is still present, it is unlikely that the primary therapy will be curative; however, therapy is continued until progressive disease demands a change.

The ongoing joint study sponsored by the Ovarian Cancer Study Group and the Gynecologic Oncology Group in early stage disease should assist the clinician in determining the role of adjuvant single-agent chemotherapy or radioisotopes. Given the poor prognosis associated with bulky residual Stage II disease, such patients might benefit from adjuvant combination chemotherapy. In advanced disease, a number of ongoing trials are attempting to evaluate the role of radiation therapy in minimal residual Stage III disease and to improve the efficacy of combination chemotherapy. This work along with research aimed at understanding etiology and prevention and speeding diagnosis may further improve the outlook of victims of this disease.

Literature Cited

1. Scully, R. E. 1977. Ovarian tumors: a review. *Am. J. Pathol.* 87:686–720
2. Steinberg, W. H., Dhurandhar, H. N. 1977. Functional ovarian tumors of stromal and sex cord origin. *Hum. Pathol.* 8:565–82
3. Stenwig, J. T., Hazecamp, J. T., Beecham, J. B. 1979. Granulosa cell tumors of the ovary: a clinicopathological study of 118 cases with long-term follow-up. *Gynecol. Oncol.* 7:136–52
4. Lingeman, C. H. 1974. Etiology of cancer of the human ovary: a review. *J. Natl. Cancer Inst.* 53:1603–18
5. Beral, V., Fraser, P., Chilvers, C. 1978. Does pregnancy protect against ovarian cancer? *Lancet* 1:1083–87
6. Schottenfeld, D., Berg, J. 1971. Incidence of multiple primary cancers. IV. Cancers of the female breast and genital organs. *J. Natl. Cancer Inst.* 46:161–70
7. King, H., Haenszel, W. 1973. Cancer mortality among foreign- and native-born Chinese in the United States. *J. Chron. Dis.* 26:623–46
8. Longo, D. L., Young, R. C. 1979. Cosmetic talc and ovarian cancer. *Lancet* 2:349–351
9. Newhouse, M. L., Berry, G., Wagner, J. L., Turok, M. E. 1972. A study of the mortality of female asbestos workers. *Br. J. Ind. Med.* 29:134–41
10. Henderson, W. J., Joslin, C. A. F., Turnbull, A. C., Griffiths, K. 1971. Talc and carcinoma of the ovary and cervix. *J. Obstet. Gynaecol. Br. Commonw.* 78:266–72
11. Casagrande, J. T., Lowe, E. W., Pike, M. C., Roy, S., Ross, R. K., Henderson, B. E. 1979. "Incessant ovulation" and ovarian cancer. *Lancet* 2:170–73
12. Parker, R. T., Parker, C. H., Wilbanks, G. T. 1970. Cancer of the ovary. *Am. J. Obstet. Gynecol.* 108:878–93
13. Feldman, G. B., Knapp, R. C. 1974. Lymphatic drainage of the peritoneal cavity and its significance in ovarian cancer. *Am. J. Obstet. Gynecol.* 119: 991–94
14. Tobias, J. S., Griffiths, C. T. 1976. Management of ovarian cancer: current concepts and future prospects. *N. Engl. J. Med.* 294:818–23, 877–82
15. Fuks, Z., Bagshaw, M. A. 1975. The rationale for curative radiotherapy for ovarian cancer. *Int. J. Radiat. Oncol. Biol. Phys.* 1:21–32
16. Rosenoff, S. H., Young, R. C., Anderson, T., Bagley, C., Chabner, B., Schein, P. S., Hubbard, S., DeVita, V. T. 1975. Peritoneoscopy: a valuable staging tool in ovarian cancer. *Ann. Intern. Med.* 83:37–41
17. DePalo, G., Musumeci, R., Kendla, R., et al. 1980. The reassessment of patients with ovarian carcinoma. *Eur. J. Cancer.* In press
18. Griffiths, C. T., Fuller, A. F. 1978. Intensive surgical and chemotherapeutic management of advanced ovarian cancer. *Surg. Clin. North Am.* 58:131–42
19. Malkasian, G. D., Decker, D. G., Webb, M. J. 1975. Histology of epithelial tumors of the ovary: clinical usefulness and prognostic significance of histologic classification and grading. *Semin. Oncol.* 2:191–201
20. Ozols, R. F., Garvin, A. J., Costa, J., Simon, R. M., Young, R. C. 1978. Histologic grade in advanced ovarian cancer. *Cancer Treat. Rep.* 63:255–63
21. Piver, S. M., Lele, S., Barlow, J. J. 1976. Preoperative and intraoperative evolution in ovarian malignancy. *Obstet. Gynecol.* 48:312–15
22. Ozols, R. F., Fisher, R. I., Anderson, T, Young, R. C. 1979. Peritoneoscopy in ovarian cancer: a safe and valuable procedure for staging and followup. *Proc. Am. Assoc. Cancer Res.* 20:408
23. Smith, J. P., Schwartz, P. E. 1980. Second-look laparotomy and prognosis related to extent of residual disease. In *Therapeutic Progress in Ovarian Cancer, Testicular Cancer and the Sarcomas,* ed. A. VanOosterom, F. M. Muggia, F. J. Cletons, pp. 77–94. The Hague: Martinus Nijhoff
24. Dunnick, N. R., Fisher, R. I., Chu, E. W., Young, R. C. 1980. Percutaneous aspiration of retroperitoneal lymph nodes in ovarian cancer. *Am. J. Roentgenol.* 135:109–13
25. Piver, M. S., Lele, S. B., Barlow, J. J., Gamana, M. 1980. Second-look laparoscopy with cytologic washings

and biopsies prior to proposed second-look laparotomy. *Obstet. Gynecol.* In press

26. Fuks, Z. 1975. External radiotherapy of ovarian cancer: standard approaches and new frontiers. *Semin. Oncol.* 2: 253–63
27. Hreschyshyn, M. M., Norris, H. G., Park, R. 1978. Postoperative treatment of resectable malignant and possibly malignant epithelial ovarian tumors with radiotherapy, melphalan, or no treatment. *Proc. XII Int. Cancer Congr.* 12:157
28. Douin, P., Rutledge, F. N., Delclos, L., Smith, J. P. 1979. Comparison of external radiotherapy and chemotherapy in ovarian cancer. *Ann. R. Coll. Physicians Surg. Can.* 12:61
29. Dembo, A. J., Bush, R. S., Beale, F. A., et al. 1979. Ovarian carcinoma: improved survival following abdominopelvic irradiation in patients with a completed pelvic operation. *Am. J. Obstet. Gynecol.* 134:793–99
30. Young, R. C., Chabner, B. A., Hubbard, S. M., et al. 1978. Advanced ovarian adenocarcinoma: a prospective clinical trial of melphalan (L-PAM) versus combination chemotherapy. *N. Engl. J. Med.* 299:1261–66
31. Edmonson, H. J., Fleming, T. R., Decker, D. G., et al. 1979. Different therapeutic sensitivities and host factors affecting prognosis in advanced ovarian carcinoma versus minimal residual disease. *Cancer Treat. Rep.* 63:241–47
32. Glatstein, E., Fuks, Z., Bagshaw, M. A. 1977. Diaphragmatic treatment in ovarian carcinoma: a new radiotherapeutic technique. *Int. J. Radiat. Oncol. Biol. Phys.* 2:357–63
33. Julian, C. J., Inalsingh, C. H. A., Burnett, L. S. 1978. Radioactive phosphorus and external radiation as an adjuvant to surgery for ovarian carcinoma. *Obstet. Gynecol.* 52:155–63
34. Julian, C. J., Woodruff, S. D. 1969. The role of chemotherapy in the treatment of primary ovarian malignancy. *Obstet. Gynecol. Surv.* 11:1307–42
35. Bush, R. S., Allt, W. E. C., Beale, F. A., et al. 1977. Treatment of epithelial carcinoma of the ovary. Surgery, irradiation, and chemotherapy. *Am. J. Obstet. Gynecol.* 127:692–704
36. Smith, J. P., Rutledge, F. N. 1970. Chemotherapy in the treatment of cancer of the ovary. *Am. J. Obstet. Gynecol.* 107:691–714
37. DeVita, V. T., Schein, P. S. 1973. The use of drugs in combination for the treatment of cancer: rationale and results. *N. Engl. J. Med.* 288:998–1006
38. Dimitriades, M., Papadimetriou, C., et al. 1978. Chemotherapy of ovarian cancer. *Int. Surg.* 63/2:81–84
39. Vogl, S. E., Berenzweig, M., Kaplan, B. H., et al. 1979. The CHAD and HAD regimens in advanced ovarian cancer: combination chemotherapy including cyclophosphamide, hexamethylmelamine, adriamycin, and *cis*-dichlorodiammine platinum (II). *Cancer Treat. Rep.* 63:311–17
40. Smith, J. P., Rutledge, F. N., Wharton, J. T. 1972. Chemotherapy of ovarian cancer: new approaches to treatment. *Cancer* 30:1565–71
41. Barlow, J. J., Piver, M. S. 1977. Single agent versus combination chemotherapy in the treatment of ovarian cancer. *Obstet. Gynecol.* 49:609–11
42. Turbow, M. M., Fuks, Z., Glatstein, E. 1978. Chemotherapy of ovarian carcinoma: randomization between melphalan and adriamycin-cyclophosphamide. *Proc. Am. Assoc. Cancer Res.* 19:394
43. Neijt, J. P., VanLindert, A. C. M., Vendrijk, C. P. J., et al. 1979. Hexa-CAF combination chemotherapy and other multiple drug regimens in advanced ovarian carcinoma: present and future. *Neth. J. Med.* 22:1–12
44. Young, R. C., VonHoff, D. D., Gormley, P., et al. 1979. *Cis*-dichlorodiammine platinum (II) for the treatment of advanced ovarian cancer. *Cancer Treat. Rep.* 63:1539–44
45. Stanhope, R. C., Smith, J. P., Rutledge, F. 1977. Second trial drugs in ovarian cancer. *Gynecol. Oncol.* 5:52–58
46. Bolis, G., D'Incalci, M., Gramellini, F., et al. 1978. Adriamycin in ovarian cancer patients resistant to cyclophosphamide. *Eur. J. Cancer* 14:1401–5
47. Hubbard, S. M., Barkes, P., Young, R. C. 1978. Adriamycin therapy for advanced ovarian carcinoma after chemotherapy. *Cancer Treat. Rep.* 62: 1375–79
48. Kardinal, C. G., Luce, J. K. 1977. Evaluation of a hexamethylmelamine and 5-fluorouracil combination in the treatment of advanced ovarian carcinoma. *Cancer Treat. Rep.* 61:1691–94
49. Briscol, K. E., Pasmantier, M. W., Ohnuma, T., et al. 1978. *Cis*-dichlorodiammine platinum (II) and adriamycin treatment of advanced ovarian cancer. *Cancer Treat. Rep.* 62:1027–31

50. Alberts, D. S., Hilgers, R. D., Moon, T. E. 1979. Combination chemotherapy for alkylator-resistant ovarian carcinoma: a preliminary report of a Southwest Oncology Group trial. *Cancer Treat. Rep.* 63:301–6
51. Kane, R., Harvey, H., Andrews, T., et al. 1979. Phase II trial of cyclophosphamide, hexamethylmelamine, adriamycin, and *cis*-dichlorodiammine platinum (II) combination chemotherapy in advanced ovarian carcinoma. *Cancer Treat. Rep.* 63:307–10
52. Bruckner, H. W., Ratner, L. H., Cohen, C. J., et al. 1978. Combination chemotherapy for ovarian cancer with cyclophosphamide, adriamycin, and *cis*-dichlorodiammine platinum (II) after failure of initial chemotherapy. *Cancer Treat. Rep.* 62:1021–26
53. Creasman, W. J., Gall, S. A., Blessing, J. A., et al. 1979. Chemoimmunotherapy in the management of primary Stage III ovarian cancer: a Gynecologic Oncology Group study. *Cancer Treat. Rep.* 63:319–24
54. Alberts, D. S., Moon, T. E., Stephens, R. A., et al. 1979. Randomized study of chemoimmunotherapy for advanced ovarian carcinoma: a preliminary report of the Southwest Oncology Group. *Cancer Treat. Rep.* 63:325–32
55. Dedrick, R. L., Myers, C. E., Bungay, P. N., DeVita, V. T. 1978. Pharmacokinetic rationale for peritoneal drug administration in the treatment of ovarian cancer. *Cancer Treat. Rep.* 62:1–11
56. Jones, R. B., Myers, C. E., Guarino, A. M., et al. 1978. High volume intraperitoneal chemotherapy ("belly bath") for ovarian cancer: Pharmacologic basis and early results. *Cancer Chemother. Pharmacol.* 1:161–65
57. Speyer, J. L., Collins, J. M., Dedrick, R. L., et al. 1980. Phase I and pharmacological studies of 5-fluorouracil administered intraperitoneally *Cancer Res.* 40:567–71
58. Ozols, R., Grotzinger, K., Myers, C., Young, R. C. 1979. Ovarian cancer: intraperitoneal therapy with adriamycin. *Proc. Am. Assoc. Cancer Res.* 20:242
59. Hamburger, A. W., Salmon, S. E. 1977. Primary bioassay of human tumor stem cells. *Science* 197:461–63
60. Salmon, S. E., Soehnlen, B., Alberts, D. S 1980. New drugs in ovarian cancer: in vitro phase II screening with the human tumor stem cell assay. See Ref. 23, pp. 113–28
61. Donahoe, P. K., Swann, D. A., et al. 1979. Mullerian duct regression in the embryo correlated with cytotoxic activity against human ovarian cancer. *Science* 205:913–15

Ann. Rev. Med. 1981. 32:491–97
Copyright © 1981 by Annual Reviews Inc. All rights reserved

CHRONIC MUCOCUTANEOUS CANDIDIASIS

❖7414

John M. Dwyer, M.D., Ph.D.[1]

Department of Medicine, Yale University School of Medicine,
New Haven, Connecticut 06510

The fungus *Candida albicans* lives within the intestinal flora of man and remains a harmless saprophyte while its proliferation is limited by intestinal bacteria, secretory IgA (1), cell-mediated immunity (2), and other nonspecific mechanisms currently poorly understood. The failure of any of these homeostatic mechanisms results in fungal overgrowth, whereupon *C. albicans* becomes a pathogen. Chronic mucocutaneous candidiasis (CMC) is a syndrome resulting from such proliferation in which the skin and nails as well as oral, esophageal, and genital mucosa become chronically infected (3). CMC is not a single disease entity but represents the final common clinical pathway resulting when one or more of the complex defense mechanisms restricting *C. albicans* proliferation is defective. At least five distinct clinical subdivisions of the CMC syndrome can be described.

Juvenile Onset Candidiasis

CMC may develop within the first few weeks or months of life. Chronic, occasionally granulomatous infection is established causing thrush, a papular diaper rash, nail dystrophy, infection of the scalp, glabrous skin, and genital mucosa (4–7). The skin lesions gradually progress into widespread erythematous lesions (4). Dermatophyte infections commonly complicate the picture (2). Fungal infections predominate, but patients often develop recurrent bacterial pneumonias and bronchiectasis, as well as infections with herpes simplex, zoster, and opportunistic organisms (4, 8, 9). Occasionally the disease occurs in families and both recessive and dominant

---

[1] Dr. Dwyer is an Investigator for the Howard Hughes Medical Institute.

0066-4219/81/0401-0491$01.00

patterns of inheritance have been described (10, 11). Patients with Type I CMC frequently have multiple defects in cell-mediated immunity (CMI). However, it should be noted that (*a*) many patients with similar immunological defects do not have CMC; and (*b*) CMI defects may be limited to the antigens of *C. albicans* and a few others since patients often have excellent CMI responses to alternate antigens. One faces a "chicken or the egg" conundrum in attempting to blame CMI defects for CMC. Chronic infection, particularly granulomatous infection, can cause immunosuppression (12) and cases of CMC have been described in which early in the disease no immunological defects exist, such changes occurring only after a number of months or even a year of infection (8). Whether defects in T cell function are primary or secondary they seem to be, at the very least, associated with perpetuation of the infection: ample evidence suggests that clinical improvement follows restoration of cell-mediated immune function to candida (see below). So far the following defects in CMI have been described in Type I CMC. (*a*) Patients are unable to produce delayed hypersensitivity to all skin test antigens or in some cases only to *C. albicans* (8). (*b*) Lymphocyte transformation to candida antigens in vitro is intact but lymphokine production and macrophage inhibition factor (MIF) are abnormal (13, 14). (*c*) Lymphocyte transformation (a T cell response) is deficient, but MIF production is normal, an apparent anomaly; but note that MIF release does not need DNA synthesis and a lymphokine may be released by B cells as well as by T cells (14, 15). (*d*) Both lymphocyte transformation and lymphokine release are abnormal (4). (*e*) A soluble inhibitor of lymphocyte transformation to *C. albicans* is present in plasma (5). (*f*) A serum inhibitory factor that prevents the killing of *C. albicans* in vitro is present (16). (*g*) Responses of monocytes (essential for delayed hypersensitivity responses to chemotactic stimuli) are impaired (17) and macrophage function is disturbed (18). In addition, abnormal complement function and IgA deficiency have been described in some patients (1).

Candidiasis Endocrinopathy Syndrome

The association of CMC with hypofunction of one or more endocrine organs is referred to as the candidiasis endocrinopathy syndrome (CES). Many reports of this syndrome document the association of a wide spectrum of endocrine and nonendocrine organ dysfunction with CMC (19, 20). In one study (21), 8 of 12 cases showed evidence of hepatitis almost certainly due to autoimmune active chronic hepatitis (19). Of 31 cases with CES (20), pernicious anemia was present in seven and alopecia totalis in six. Other nonendocrine findings reported include pulmonary fibrosis (21), enamel hypoplasia, and keratoconjunctivitis (22). Hypoparathyroidism is the most common endocrine defect followed by adrenal failure (22). Other associations reported include chronic lymphocytic thyroiditis (23), diabetes mellitus, ovarian failure, and ACTH deficiency (24). CES can also be familial (19, 25). In a recent family study (19) a number of clinically affected and nonaffected members had demonstrable abnormalities in immunoregulatory T cells. The disordered immunoregulation in CES may explain the autoimmune disease, absent IgA and elevated IgE levels, polyclonal hypergammaglobulinemia, and pronounced failure of lymphocytes to respond to in vitro challenge with candida antigens found in many members of this family. The most difficult aspect of this disease to explain, as in all CMC, remained the candidiasis itself. In CES as elsewhere it appears that an effective host response to this fungus requires optimal integrity and cooperation among the various participants of the immune response.

Later Onset

The forms of CMC mentioned above usually occur early in life. There is another form of candidiasis that occurs during the end of the first or early in the second decade and is usually not associated with a history of susceptibility to infectious disease during infancy. Such patients tend to have milder disease and often, after treatment with antifungal therapy, spontaneously develop CMI to *C. albicans* (8).

CMC and Thymoma

Another group of patients with CMC are apparently healthy during childhood and young adult life, but during or after the fourth decade they develop CMC. The infection is very resistant to antifungal agents and the group is characterized by the presence, in the majority of patients, of a thymoma. CMI to *C. albicans* is often suppressed and antinuclear and other autoantibodies are frequently present. Thymic hormone levels are normal in plasma but plasma inhibitors of the response to candida have been found in some patients. Tumors are rare outside the thymus but neoplastic changes can occur within the gland itself. While the syndrome is not associated with endocrine disorders, it has been reported frequently in association with myasthenia gravis, myositis, alopecia areata, pernicious anemia, hypogammaglobulinemia, vitiligo, and lymphocytic leukemia (26).

CMC without Associated Abnormalities

Perhaps 25–30% of patients with CMC have no immunological defects. Despite apparently normal CMI responses to candida these patients appear to have little defense against this organism. Such cases may be milder than those with clear-cut immunological defects but the author has seen severe disease including destructive esophageal lesions in this group. It should be noted that, in contradistinction to the patient with generalized immunodeficiency, systemic candidiasis does not occur in any of the conditions described above.

Treatment

CMC remains incurable but currently available therapy significantly reduces the morbidity suffered by most patients. Topical therapy with the nystatin or imidazole group of drugs is useful adjunct therapy (8). Amphotericin B is generally very effective but the well-known side-effects limit its application. Iron deficiency is a constant finding in CMC even in patients with a normal hemoglobin. In patients with delayed onset CMC, restoration of normal iron levels effectively minimizes the clinical manifestations of the disease and even normalizes immune responses (3). In 50–75% of patients with defects in CMI (generalized or candida specific) transfer factor (TF), a low molecular weight dialyzable substance obtained from normal leukocytes, improves the clinical condition (27, 28). TF must be obtained from donors who exhibit strong delayed hypersensitivity reactions to *Candida albicans* (30). TF is most effective when the antigen load is first reduced by amphotericin B and a synergistic effect was reported when TF was administered along with fetal thymic tissue (30). The thymic hormone, thymosin, does not improve CMC (31), a fact that correlates with earlier reports of normal plasma thymic hormone levels in such patients (32).

Increased attention is being paid to esophageal infection in all forms of CMC (33). Severe esophagitis with burning, dysphagia, and nausea are common. Mucosal infiltration sets up an inflammatory response probably activated by the alternate pathway of complement and results in problems ranging from discomfort to a stricture severe enough to require dilatation (11). Systemic treatment of CMC will assist the esophagitis but local therapy may also be useful. Gargling and swallowing an amphotericin solution six times per day is of proven value (34). Topically applied amphotericin cream (3%) may assist vulvovaginal and perirectal infection. Miconazole administered intravenously has helped some patients but initial enthusiasm for this drug has waned (35). The drug requires an extended period of administration but is much less toxic than amphotericin B. Clotrimazole, useful topically, is very effective when taken orally but gastrointestinal side-effects prevent all but a few patients from taking the drug.

Clotrimazole troches have been used more successfully (36). Ten of ten recipients experienced a marked regression of symptoms and mucosal lesions when using 5 X 10 mg troches per day. Of the drugs currently under investigation, the most promising appears to be the imidazole derivative ketoconazole. At an international symposium on this drug held in Colombia in 1979, it was reported[2] that of 30 patients with CMC treated with ketoconazole, 17% were moderately improved, 40% markedly improved, and 40% apparently cured. Two other potential advances require further clinical study. Levamisole, also an imidazole derivative, nonspecifically enhances various parameters of the immune response. Though not without side-effects (severe granulocytopenia has been observed), neutrophil chemotactic defects in many patients including those with CMC have been improved by this drug (37). Another nonspecific enhancer of cell-mediated immunity is the H-2 histamine receptor antagonist, cimetidine. Suppressor cells carry surface H-2 histamine receptors, and antagonists may enhance the ability of effector cells to produce strong cell-mediated immunity by depressing suppressor cells (38). In four patients with CMC, a four-week course of cimetidine (300 mg qid) produced good delayed hypersensitivity and lymphokine production after stimulation with candida antigens (39). No demonstrable clinical improvement occurred with this short course of therapy and long-term studies are needed to prove its effectiveness.

The pathogenesis of CMC no doubt holds many lessons for the immunologist, endocrinologist, and dermatologist. The ultimate clinical advances, however, are unlikely to be made until we can answer that most fundamental and still most puzzling of relevant questions: how is *Candida albicans* usually kept in its place?

---

[2] J. Symoens, M. Moens, J. Dom, H. Scheijgrond, J. Dony, V. Schuermans, R. Legendre, N. Finestine.

Literature Cited

1. Chilgren, R. A., Quie, P. G., Meuwissen, H. J. 1967. Chronic mucocutaneous candidiasis deficiency of delayed hypersensitivity, and selective local antibody defect. *Lancet* 2:688–93
2. Valdimarsson, H., Higgs, J. M., Wells, R. S., Yamamura, M., Hobbs, J. R., Holg, P. J. L. 1973. Immune abnormalities associated with chronic mucocutaneous candidiasis. *Cell. Immunol.* 6:348–61
3. Rosman, N. 1979. Chronic mucocutaneous candidiasis. *Postgrad. Med. J.* 55:611–14
4. Kirkpatrick, C. H., Montes, L. F. 1974. Chronic muco-cutaneous candidiasis. *J. Cutan. Pathol.* 1:211–29
5. Canales, L., Middlemas, R. O., Louro, J. M. 1969. Immunological observations in chronic muco-cutaneous candidiasis. *Lancet* 2:567–70
6. Edwards, J. E., Lehrer, R. I., Stiehm, E. R., Fischer, T. J., Young, L. S. 1978. Severe candidal infections. Clinical perspective, immune defense mechanisms, and current concepts of therapy. *Ann. Intern. Med.* 89:91–106
7. Chilgren, R. A., Meuewissen, H. J., Quie, P. G., Good, R. A., Hong, R. 1969. The cellular immune defect in chronic mucocutaneous candidiasis. *Lancet* 2:1286–88
8. Sams, W. M., Jorizzo, J. L., Snyderman, R., Jegasothy, B. V., Ward, F. E., Weiner, M., Wilson, J. G., Young, W. J., Dillard, S. B. 1979. Chronic mucocutaneous candidiasis. Immunologic studies of three generations of a single family. *Am. J. Med.* 67:948–59
9. Chipps, B. E., Saulsbury, F. T., Hsu, S. H., Hughes, W. T., Winkelstein, J. A. 1979. Non-candidal infections in children with chronic mucocutaneous candidiasis. *Johns Hopkins Med. J.* 144: 175–79
10. Wuepper, K. D., Fudenberg, H. H. 1967. Moniliasis, "auto-immune" polyendocrinopathy, and immunologic family study. *Clin. Exp. Immunol.* 2:71–82
11. Kirkpatrick, C. H., Rich, R. R., Bennett, J. E. 1971. Chronic mucocutaneous candidiasis. Model-building in cellular immunity. *Ann. Intern. Med.* 74:955–78
12. Dwyer, J. M., Mangi, R. J., Gee, B., Kantor, F. S. 1976. Comparison of the anergy of sarcoidosis with experimentally induced anergy in guinea pigs. *Ann. NY Acad. Sci.* 278:29–35
13. Lehner, T., Wilton, J. M. A., Ivanyi, L. 1972. Immunodeficiencies in chronic mucocutaneous candidosis. *Immunology* 22:775–87
14. VanScoy, R. E., Hill, H. R., Ritts, R. E. Jr., Quie, P. G. 1975. Familial neutrophil chemotaxis defect, recurrent bacterial infections, mucocutaneous candidiasis, and hyperimmunoglobulinemia E. *Ann. Intern. Med.* 82:766–71
15. Goldberg, L. S., Bluestone, R., Barnett, E. V., Landau, J. W. 1971. Studies on lymphocyte and monocyte function in chronic mucocutaneous candidiasis. *Clin. Exp. Immunol.* 8:37–43
16. Chilgren, R. A., Oh, M., Hong, R. 1969. Inhibition of polymorphonuclear killing of *Candida albicans. Fed. Proc.* 28:498
17. Synderman, R., Altman, L. C., Frankel, A. 1973. Defective mononuclear leukocyte chemotaxis. A previously unrecognized immune dysfunction. *Ann. Intern. Med.* 78:509–13
18. Quie, P. G., Chilgren, R. A. 1971. Acute disseminated and chronic mucocutaneous candidiasis. *Semin Hematol.* 8:227–42
19. Arulanantham, K., Dwyer, J. M., Genel, M. 1979. Defective immunoregulation in familial candidiasis endocrinopathy syndrome. *N. Engl. J. Med.* 300:164–68
20. Blizzard, R. M., Gibbs, J. H. 1968. Studies pertaining to its association with endocrinopathies and pernicious anemia. *Pediatrics* 42:231–37
21. Whitaker, J., Landing, B. H., Esselborn, V. M., Williams, R. R. 1956. Syndrome of familial juvenile hypoadrenocorticism, hypoparathyroidism and superficial moniliasis. *J. Clin. Endocrinol.* 16:1374–87
22. Hermans, P. E., Ulrich, J. A., Markowitz, H. 1969. Chronic mucocutaneous candidiasis as a surface expression of deep seated abnormalities. *Am. J. Med.* 47:503–19
23. Kenny, F. M.. Holliday, M. A. 1964. Hypoparathyroidism, moniliasis, Addison's disease and Hashimoto's disease. Hypercalcemia treated with intravenously administered sodium sulphate. *N. Engl. J. Med.* 271:708–13
24. Castells, S., Fikrig, S., Inamdar, S., Orti, E. 1971. Familial moniliasis, defective delayed hypersensitivity and adrenocorticotropic hormone deficiency. *J. Pediatr.* 79:72–79

25. Kroll, J. J., Einbinder, J. M., Merz, W. G. 1973. Muco-cutaneous candidiasis in a mother and son. *Arch. Dermatol.* 108:259–61
26. Kirkpatrick, C. H., Windhorst, D. B. 1979. Mucocutaneous candidiasis and thymoma. *Am. J. Med.* 939–45
27. Shulkind, M. L., Adler, W. H., Altemeir, W. A. III, Ayoub, E. M. 1972. Transfer factor in the treatment of a case of chronic mucocutaneous candidiasis. *Cell. Immunol.* 3:606–15
28. Fudenberg, H. H., Spitler, L. E., Levin, A. S. 1972. Treatment of immune deficiency. *Am. J. Pathol.* 69:529–35
29. Littman, B. H., Rocklin, R. E., Parkman, R., David, J. R. 1978. Transfer factor treatment of chronic mucocutaneous candidiasis: requirement for donor reactivity to candida antigen. *Clin. Immunol. Immunopathol.* 9:97–110
30. Ballow, M., Hyman, L. W. 1977. Combination immunotherapy in chronic mucocutaneous candidiasis synergism between transfer factor and fetal thymus tissue. *Clin. Immunol. Immunopathol.* 8:504–12
31. Akhter, J., Merker, R. I., Rieger, C. H. L., Rothberg, R. M. 1980. Effect of thymosin on lymphocytes from patients with chronic mucocutaneous candidiasis and endocrinopathies *J. Allergy Clin. Immunol.* 65:34–40
32. Kirkpatrick, C. H., Greenberg, L. E., Chapman, S. W., Goldstein, G., Lewis, V. M., Twomey, J. J. 1978. Plasma thymic hormone activity in patients with chronic mucocutaneous candidiasis. *Clin. Exp. Immunol.* 34:311–17
33. Jensen, K. B., Stenderup, A., Thomsen, J. B. 1964. Esophageal moniliasis in malignant neoplastic disease. *Acta Med. Scand.* 175:715–19
34. Ezdinli, E. Z., O'Sullivan, D. D., Wasser, L. P. 1979. Oral amphotericin candidiasis in patients with hematologic neoplasms; an autopsy study. *J. Am. Med. Assoc.* 142:258–61
35. Stevens, D. A. 1977. Miconazole in the treatment of systemic fungal infections. *Am. Rev. Respir. Dis.* 116:801–3
36. Kirkpatrick, C. H., Alling, D. W. 1978. Treatment of chronic oral candidiasis with clotrimazole troches. A controlled clinical trial. *N. Engl. J. Med.* 299:1201–3
37. Hogan, N. A., Hill, H. R. 1978. Enhancement of neutrophil chemotaxis and alteration of levels of cellular cyclic nucleotides by levamisole. *J. Infect. Dis.* 138:4:437–44
38. Rocklin, R. E. 1976. Modulation of cellular immune responses in vivo and in vitro by histamine receptor bearing lymphocytes. *J. Clin. Invest.* 57:1051–58
39. Jorizzo, J. L., Sams, W. M. Jr., Jegasothy, B. V., Olansky, A. J. 1980. Cimetidine as an immunomodulator: chronic mucocutaneous candidiasis as a model. *Ann. Intern. Med.* 92:192–95

Ann. Rev. Med. 1981. 32:499–507
Copyright © 1981 by Annual Reviews Inc. All rights reserved

PROGNOSIS IN BUNDLE BRANCH BLOCK

❖7415

J. McAnulty, M.D. and S. Rahimtoola, M.D.

University of Oregon Health Sciences Center, Portland, Oregon 97201

Once considered likely to progress to complete heart block, bundle branch block was regarded as an ominous finding on the electrocardiogram. Closer evaluation has shown that progression to symptomatic bradyarrhythmias is uncommon (1–7). However, some may develop symptoms suggestive of an intermittent bradyarrhythmia, particularly dizziness or syncope, and, on occasion, this can be shown to be caused by complete heart block (Figure 1). This rhythm is associated with significant morbidity and mortality and is one of many reasons why it is important to understand the clinical significance of bundle branch block.

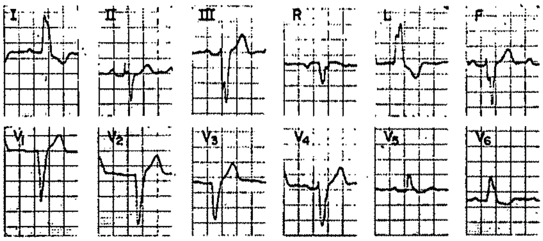
PREVALENCE

Bundle branch block is common and its incidence and prevalence increase with age. It was found in 0.2% of a large population of US Air Force personnel who averaged under 40 years of age (8), in 0.6% of those who entered the Framingham Study (the age of those affected averaged 48 and 49 years for right and left bundle branch block respectively) (6, 7), and in 1–2% of large patient populations averaging over 60 years (1–3, 6, 7). It is an electrocardiographic finding that is and will be encountered frequently.

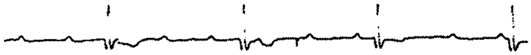
ETIOLOGY

Bundle branch block is a marker for underlying heart disease. Clinically apparent organic heart disease is recognizable, and the likely cause of the conduction abnormality, in 80% of those with bundle branch block (1–7). Coronary artery disease is recognized in almost 50% of these patients, half of whom have had a myocardial infarction (2). Hypertension is found in up

0066-4219/81/0401-0499$01.00



1973 ECG



1977 RHYTHM STRIP WHEN ADMITTED FOR SYNCOPE

Figure 1 The electrocardiogram from a patient who presented with syncope in 1973 revealed left bundle branch block and left axis deviation (*top*). Unlike many with bundle branch block, she had recurrent episodes of syncope. Despite multiple attempts monitoring did not reveal the cause of her symptoms until 1977 when complete heart block was demonstrated (*bottom*). A permanent pacemaker was inserted.

to 50% and valvular heart disease, cardiomyopathies, or congenital heart disease are present in 10–20% (1, 2). In 20% of patients with bundle branch block, there is no evidence of clinical cardiovascular disease (1–3). It is likely that most of these individuals have a disease process involving the conduction system alone: primary conduction system disease. However, in some, associated cardiac disease may simply not be recognized because patients with bundle branch block are more likely to develop recognizable coronary artery disease and congestive heart failure than patients without bundle branch block (7, 8).

ANATOMY

Present concepts about the anatomy of the conduction system support the clinical concern that bundle branch block is likely to progress to complete heart block. Though the distal conduction system is more likely a syncitium, the assumption that it is trifascicular has some validity (Figure 2). In an unaffected individual at least three distal pathways connect the AV node to the ventricle, and block of any one of these is recognizable on the electrocardiogram (Tables 1 and 2). If the patient has block of one of these three fascicles, *unifascicular block,* there are still two others capable of transmitting the electrical impulse to the ventricle. However, when there is electrocardiographic evidence of block of two fascicles, *bifascicular block,* the potential for complete heart block increases: Only one pathway remains open for conduction and it, too, could be affected by the same disease

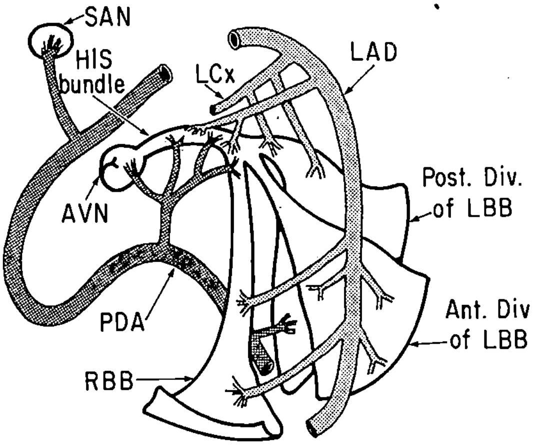


Figure 2 This diagrammatic representation of the conduction system and its blood supply demonstrates the trifascicular nature of the distal conduction system with the three fascicles being the right bundle, the anterior division of the left bundle and the posterior division of the left bundle. SAN: sinus node; LBB: left bundle branch; RBB: right bundle branch; AVN: atrioventricular node; PDA: posterior descending artery; LAD: left anterior descending artery; LCA: left circumflex artery.

process that caused block in the other two. In fact, intracardiac electrophysiologic studies have revealed disease in the third fascicle in two thirds of patients with bundle branch block (2). Patients with bifascicular block do have an increased incidence of complete heart block (2, 4) and thus some clinical concern based on this anatomical concept is reasonable.

Clinical Events

Bundle branch block is associated with syncope, complete heart block, death, and sudden death. An increasing amount of data is accumulating about its contribution to each.

SYNCOPE When patients are first seen with bundle branch block, up to 30% have already had syncope (2), that symptom which would seem most alarming when considering the possibility of complete heart block. On occasion, this bradyarrhythmia may be documented as the cause, but more often it is not clear that a bradyarrhythmia caused the symptoms. Frequently, syncope can be shown to be due to causes other than complete heart block, such as tachyarrhythmias or primary neurologic problems (2, 9). Even though an alarming symptom, syncope is likely to recur in less than 50% of bundle branch block patients as judged by a study with a follow-up period of over four years (2). Dizziness also can be caused by transient heart block, but it is an even less easily interpretable symptom and there is no

Table 1 Electrocardiographic criteria of intraventricular conduction defects

Right bundle branch block (RBBB)	QRS duration ≥ 0.12 seconds RSR′ or notched R wave or QR (if previous MI) with ST depression and T inversion in V_1 Terminal S wave in leads I and V_6
Left bundle branch block (LBBB)	QRS duration ≥ 0.12 seconds Presence of a monophasic notched R wave in leads I and V_6 ST depression, T inversion in V_6 Absence of Q and S wave in V_6
Atypical LBBB	Same criteria as for LBBB except that a Q or S wave is present in V_6
Left axis deviation (LAD)	Mean QRS axis in the frontal plane $\leq -30°$
Left anterior fascicular block (left anterior hemiblock)	Left axis deviation with initial R waves in II, III, AVF; Q waves in I, AVL
Right axis deviation (RAD)	Mean QRS axis with frontal plane $\geq 105°$
Left posterior fascicular block (left posterior hemiblock)	Right axis deviation with Q waves in II, III, AVF and initial R waves in I, AVL in patients without right ventricular hypertrophy
Prolonged PR interval (first-degree AV block)	PR interval > 0.20 seconds

conclusive evidence that it is commonly caused by transient complete heart block.

COMPLETE HEART BLOCK Chronic bundle branch block can progress to complete heart block (Figure 1), but this probably occurs in less than 2% of patients with chronic bundle branch block per year (2, 4) with an incidence exceeding 5% only in those with severe congestive heart failure (10). Because bundle branch block is a distal conduction system disease, heart block would be assumed to occur below the His bundle. However, in over half the cases it likely occurs in the AV node (4), which indicates that those with bundle branch block also have disease elsewhere in the conduction system.

SURVIVAL Individuals with bundle branch block have an increased mortality rate when compared to patients without this conduction abnormality. Most have associated heart disease that is the cause of death (1, 2, 5). Still, by progressing to complete heart block, bundle branch block can contribute to the mortality. Sudden death is of greatest concern when considering the contribution of the bundle branch block because death due to intermittent complete heart block, and potentially preventable, would most likely occur

Table 2 Diagnostic criteria based on surface electrocardiogram evidence[a]

Unifascicular disease[b]	Bifascicular disease[b]	Trifascicular disease
Isolated RBBB, LAFB, or LPFB	LBBB RBBB and LAFB RBBB and LPFB Alternating LAFB and LPFB	Alternating RBBB and LBBB RBBB with alternating LAFB and LPFB

[a] RBBB = right bundle branch block; LAFB = left anterior fascicular block; LPFB = left posterior fascicular block; LBBB = left bundle branch block.

[b] Trifascicular disease cannot be excluded in these patients without a measurement of an HV interval. If the HV interval is normal, trifascicular disease as manifest by conduction delay is excluded and if the HV interval is prolonged, trifascicular disease is likely to be present. The PR interval does not differentiate bifascicular from trifascicular disease.

rapidly. In this patient population, sudden death occurs at a rate of 1–10% per year in those with left bundle branch block and 1–5% per year in those with right bundle branch block and axis deviation (1, 2, 6, 7). Death due to documented complete heart block is rare, occurring in less than 0.2% of bundle branch block patients per year; most sudden deaths can be shown to be due to causes other than heart block (1, 2). In some cases the cause of sudden death cannot be determined and heart block could possibly have contributed. However, this, too, is uncommon and it appears that, at most, less than 1% of patients per year with chronic bundle branch block could be dying from complete heart block (1, 2, 11, 12).

Identification of Those at High Risk

The chance of death related to heart block in patients with bundle branch block is small; nevertheless efforts have been made to identify those who are at particularly high risk. The electrocardiogram has been evaluated, and though it was initially felt that patients with left bundle branch block were at greater risk than those with right bundle branch block and axis deviation, this has not been a universal finding (1, 2). No particular electrocardiographic abnormalities predict those who are at increased risk of developing heart block. Evaluating the PR interval in conjunction with the bundle branch block has not helped define those who are more likely to develop bradyarrhythmias (1, 2). Symptoms as a predictor of those patients with bundle branch block who are likely to develop heart block and death have been evaluated as well. Syncope has been most thoroughly studied. The patient with bundle branch block who gives a history of syncope is not clearly at increased risk of death or of sudden death (1, 2, 9). If it is assumed that a patient's syncope is due to complete heart block and yet this is not documented, insertion of a pacemaker is unlikely to prolong life or even to result in improvement of symptoms (5). Other symptoms such as dizziness, chest pain, or congestive heart failure are associated with decreased survival but do not predict which patient is at risk of death from heart block. Etiology of the bundle branch block has also been evaluated as a predictor of those patients who are likely to develop heart block or sudden death. Though those with primary conduction system disease may have a better prognosis than those with recognizable associated cardiac disease (12), etiology is not able to predict which patient will develop heart block.

It is hoped that the data obtained from an intracardiac electrophysiologic study will aid in predicting which patients are at high risk from their bundle branch block. Because it is an indicator of distal conduction, the HV (or HQ) interval may be a potentially useful prognostic indicator but by itself it may be of limited value (1, 2, 10, 14, 15). HV prolongation (> 55 msec) is associated with more severe heart disease than is a normal HV interval (15) but has not yet been shown to be an independent risk factor for complete heart block or sudden death. The greatly prolonged HV interval (> 70 msec) has been associated in some studies with increased sudden death and heart block (2, 10) but this has not been a universal finding (14). The ability to demonstrate block below the His bundle with atrial pacing may be useful in defining those who are likely to develop complete heart block (16). The exact role of intracardiac electrophysiologic data in predicting who is at risk from their bundle branch block is yet to be fully resolved.

MANAGEMENT OF CHRONIC BUNDLE BRANCH BLOCK

It is uncommon for patients with chronic bundle branch block to develop complete heart block or sudden death due to bradyarrhythmias. Because of this and because the definitive treatment for protection against bradyarrhythmias, a permanent pacemaker, is associated with significant morbidity, expense, and even mortality, a conservative approach to the management of these patients is recommended. Though chronic bundle branch block should be considered a marker for underlying heart disease, in the asymptomatic patients no further evaluation or treatment of the conduction status is required. In patients with bundle branch block who develop documented bradyarrhythmias, a permanent pacemaker should be inserted.

It is more difficult to manage patients with bundle branch block who have symptoms that could possibly be caused by bradyarrhythmias but in whom a bradyarrhythmia cannot be documented. Though no one particular diagnostic study is of high yield in defining the cause of symptoms in this group, evaluation is appropriate. Noncardiac causes of transient neurologic episodes such as seizure disorders and metabolic or vascular abnormalities should be considered, but, if none is readily apparent, the presence of bundle branch block on the electrocardiogram should increase the suspicion that a transient bradyarrhythmia was the cause of symptoms. Each patient should undergo ambulatory electrocardiographic monitoring for at least 48 hours and for longer if symptoms are recurrent. Because symptoms are frequently intermittent however, the study is often nondiagnostic. An exercise stress test may demonstrate a propensity for inappropriate bradyarrhythmias, particularly in people who have exertionally related dizziness or syncope or who may demonstrate exertional hypotension or tachyarrhythmias as the cause of symptoms. Though the intracardiac electrophysiologic study results have limitations, in the patient with no other explanation for recurrent and serious symptoms, the demonstration of a greatly prolonged HV interval or block below the His bundle with atrial pacing may be helpful. This study may also demonstrate disease of the atrioventricular or sinus node or may reveal a propensity for tachyarrhythmias, which may be helpful in deciding the need for treatment. If a bradyarrhythmia cannot be documented as the cause of symptoms in this patient group, a permanent pacemaker should not be implanted for the purpose of preventing symptoms or prolonging life (2, 5). Occasionally, an individual is in an occupation or a situation where syncopal episodes are particularly hazardous and an attempt at treatment with a permanent pacemaker may be considered.

Managing Other Medical Conditions in Bundle Branch Block

Patients with bundle branch block often have other medical problems and the conduction system status should be considered when they are treated. For example, when a patient with bundle branch block undergoes a general surgical procedure it is appropriate to ask whether they are at increased risk of developing heart block. Available evidence would suggest that this is uncommon, occuring in less than 2% of such patients during surgery (17). It is not necessary to insert a temporary pacemaker in all bundle branch block patients undergoing general surgery, but continuous rhythm monitoring in the perioperative period and for at least 24 hours after surgery is indicated. Insertion of a large central venous line prior to surgery is appropriate, to facilitate insertion of a pacing wire if necessary.

Many patients with bundle branch block have atrial and ventricular tachyarrhythmias. Occasionally, antiarrhythmic agents such as quinidine, procainamide, and disopyramide are required. These are known to depress distal conduction and prolong the HV interval, so their use demands close clinical observation for bradyarrhythmias at the start of therapy. This is less likely to occur with lidocaine and diphenylhydantoin though in large boluses they, too, can depress conduction.

Patients with bundle branch block frequently have associated heart disease for which they may need to undergo diagnostic cardiac catheterization. Insertion of a catheter in the right ventricle may induce transient right bundle branch block. In the patients with left bundle branch block prior to the study, this may result in complete heart block. The same potential hazard exists for patients with right bundle branch block who undergo catheterization of the left side of the heart. It does not seem necessary to insert a temporary pacemaker wire in all these patients prior to the procedure but careful monitoring and access to pacemaker insertion is imperative.

SUMMARY

Chronic bundle branch block is common and is a marker for associated heart disease. In the asymptomatic patient no diagnostic studies or treatment are required. In the patient with symptoms suggestive of a bradyarrhythmia, attempts to document complete heart block should be made and if this rhythm is demonstrated a permanent pacemaker should be inserted. If no arrhythmia is documented, close clinical follow-up is warranted but insertion of a permanent pacemaker has not been shown to be useful in preventing symptoms or prolonging life.

Literature Cited

1. Denes, P., Dhingra, R. C., Wu, D., Wyndham, C. R., Amat-y-Leon, F., Rosen, K. M. 1977. Sudden death in patients with chronic bifascicular block. *Arch. Int. Med.* 137:1005
2. McAnulty, J. H., Rahimtoola, S. H., Murphy, E. S., Kauffman, S., Ritzmann, L., Kanarek, P., DeMots, H. 1978. A prospective study of sudden death in "high-risk" bundle branch block. *N. Engl. J. Med.* 299:209
3. McAnulty, J. H., Kauffman, S., Murphy, E., Kassebaum, D. G., Rahimtoola, S. H. 1978. Survival in patients with intraventricular conduction defects. *Arch. Int. Med.* 138:30
4. Dhingra, R. C., Wyndham, C., Amat-y-Leon, F., Denes, P., Wu, D., Sridher, S., Bustin, A. G., Rosen, K. M. 1979. Incidence and site of atrioventricular block in patients with chronic bifascicular block. *Circulation* 59:238
5. Peters, R. W., Scheinman, M. M., Modin, G., O'Young, J., Somelefski, C. A., Mies, C. 1979. Prophylactic permanent pacemakers for patients with chronic bundle branch block. *Am. J. Med.* 66:978
6. Schneider, J. F., Thomas, H. E. Jr., Kreger, B. E., McNamara, P. M. Kannel, W. B. 1979. Newly acquired left bundle branch block: the Framingham Study. *Ann. Int. Med.* 90:303
7. Schneider, J. F., Thomas, H. E. Jr., Kreger, B. E., McNamara, P. M., Sorlie, P., Kannel, W. B. 1980. Newly acquired right bundle branch block: the Framingham Study. *Ann. Int. Med.* 92:37
8. Rotman, M., Triebwasser, J. H. 1975. A clinical and follow-up study of right and left bundle branch block. *Circulation* 51:477
9. Dhingra, R. C., Denes, P., Wu, D. Chuquimia, R., Amat-y-Leon, F., Wyndham, C., Rosen, K. M. 1974. Syncope in patients with chronic bifascicular block: significance causative mechanisms, and clinical implications. *Ann. Int. Med.* 81:302
10. Scheinman, M. D., Peters, R. W., Modin, G., Brennan, M., Mies, C.,

O'Young, J. 1977. Prognostic value of infranodal conduction time in patients with chronic bundle branch block. *Circulation* 56:240

11. Surawicz, B. 1979. Editorial: prognosis of patients with chronic bifascicular block. *Circulation* 60:40
12. Dhingra, R. C., Wyndham, C., Bauernfeind, R., Denes, P., Wu, D., Swiryn, S., Rosen, K. M. 1979. Significance of chronic bifascicular block without apparent organic heart disease. *Circulation* 60:33
13. Deleted in proof
14. Dhingra, R. C., Denes, P., Wu, D., Wyndham, C., Amat-y-Leon, F., Towne, W., Rosen, K. 1976. Prospective observations in patients with chronic bundle branch block and marked HV prolongation. *Circulation* 53:600
15. Rosen, K. M., Dhingra, R. C., Wyndham, C. R. C., Deedwania, P., Palileo, E., Bauernfeind, R. A., Swiryn, S. 1980. Significance of HV interval in 515 patients with chronic bifascicular block. *Am. J. Cardiol.* 45:405 (Abstr.)
16. Dhingra, R. C., Wyndham, C., Bauernfeind, R., Swiryn, S., Deedwania, P. C., Smith, T., Denes, P., Rosen, K. M. 1979. Significance of block distal to the His bundle induced by atrial pacing in patients with chronic bifascicular block. *Circulation* 60:1455
17. Pastore, J. O., Yurchak, P. M., Janis, K. M., Murphy, J. D., Zir, L. M. 1978. The risk of advanced heart block in surgical patients with right bundle branch block and left axis deviation. *Circulation* 57:677

Ann. Rev. Med. 1981. 32:509–20
Copyright © 1981 by Annual Reviews Inc. All rights reserved

MALE ERECTILE DISORDERS

❖7416

Raul C. Schiavi, M.D.

Human Sexuality Program, Department of Psychiatry, Mount Sinai School
of Medicine, New York, New York 10029

INTRODUCTION

Erectile impotence is defined as the inability to attain or maintain a penile erection until completion of the sexual act. Erectile disorders are etiologically classified as "psychogenic" or "organic," and it has been repeatedly stated, without supporting evidence, that less than 10% of the erectile disorders are organic. This neat dichotomy has didactic value, but it obscures the notion that in many patients organic and psychological factors frequently coexist.

There is no information about the current prevalence of erectile impotence in the population at large. In Kinsey's series (1) the prevalence was less than 2% until the age of 40; it increased to 6.7% at age 55, and by age 70 more than one quarter of the males reported long-term erectile incapacity. It is not certain the extent to which the increased prevalence of impotence with age is indicative of psychosocial factors, covert medical conditions, or the natural biological changes induced by aging.

This article briefly reviews the physiology of male sexual function as a basis for discussing recent information on disturbances in erectile capacity. The emphasis on the pathophysiological aspects of erectile impotence should not obscure the fact that, for most patients, psychological factors play a critical role in the development and maintenance of erectile difficulties.

PHYSIOLOGY OF MALE SEXUAL FUNCTION

Normal male sexual function depends upon the integrity of mediating neural, vascular, and hormonal processes. At present the neurovascular mechanisms that contribute to erection remain little understood. Animal

0066-4219/81/0401-0509$01.00

experimentation and studies of patients with spinal cord injuries suggest the existence of two distinct but interacting neurophysiologic pathways. Reflex erections are thought to occur by activation of a sacral arc via afferent impulses carried by pudendal and pelvic nerves from nervous receptors in the genitalia and from enteroceptive stimulation from the bladder or rectum. The efferent component, considered to be parasympathetic, arises from sacral segments 2–4 and reaches genital structures and the urinary bladder through the pelvic nerves (nervi erigentes). There is evidence that psychologically induced erections are mediated by a thoracolumbar sympathetic outflow that originates primarily from T12-L1 and is carried by cholinergic fibers traveling to genital structures with the hypogastric nerves (2, 3).

Only recently have the neurophysiologic and neurochemical processes that influence sexual function begun to be unravelled. Neurophysiologic studies have shown that structures within or closely related to the limbic system influence sexual function (4, 5). Brain monoaminergic and serotonergic systems control hypothalamic-pituitary activity and participate in the mediation of the central effects of sex hormones on brain processes. It has been postulated, mainly from animal research, that drugs which diminish brain dopaminergic activity or enhance serotonergic function inhibit sexual behavior (6). However, methodological and theoretical considerations render much of this literature difficult to extrapolate to human behavior (7).

The blood supply to the erectile tissues contained in the corpora cavernosa and corpus spongiosum is provided by the terminal branches of the internal pudendal arteries. Under flaccid conditions the anastomosis between the arterioles and the erectile bodies are thought to be closed by pads (polsters) first described by Ebner. Conti (8), in a classic histological study, postulated that erections occur when relaxation of the circular and longitudinal fibers of the pads, which are under autonomic nervous control, results in an increased rate of arterial inflow into the penis. Polsters located in drainage veins from the penile cavernous bodies (8, 9) and reflex contraction of the ischiocavernosus and bulbocavernosus (10) muscles may also contribute to maintain the erection by obstructing penile venous outflow. These notions have not found support in a recent isotopic penile blood outflow investigation (11) nor in a study of patients with flaccid paralysis of the perineal musculature (12).

It has been traditionally thought that autonomic influences on erectile activity involve cholinergic mechanisms. However, the observation that, in animals and humans alike, atropine does not prevent erections suggests the existence of noncholinergic vasodilator nervous processes (13). Recently, extensive adrenergic innervation has been found in the corpora cavernosa of various species including men (14). Increased sympathetic activity may result in the contraction of smooth muscle within the cavernous tissues, thereby decreasing blood flow resistance (15, 16). In addition, sympathetic constriction of arteries that supply nonerectile tissues in the penis may also help sustain tumescence. Pharmacological studies also suggest the participation of the adrenergic nervous system. Erections are enhanced by $\beta 2$ adrenergic stimulants as well as by α-adrenergic blocking agents (17).

The emission and ejaculation phases of the male sexual response are considered only briefly in this review. Emission of the sperm into the bulbar urethra is primarily controlled by the thoracolumbar sympathetic outflow via hypogastric innervation of the prostate gland, ductus deferens, and seminal vesicles. Adrenergic activation results in the rhythmic contraction of the smooth muscles of these organs, which propels the semen into the proximal urethra, and also in the closure of the neck of the urinary bladder, which prevents semen reflux. Recent pharmacological evidence suggests that the emission phase is mediated by release of norepinephrine acting on α-adrenergic receptors and that it may also be influenced by cholinergic nervous activity. Ejaculation is a reflex mediated by the pudendal nerves which results in clonic contractions of the striated muscles that surround the bulbar urethra with subsequent ejection of semen (3).

A sufficient amount of circulating androgen is required to permit adequate male sexual function. Normal adult men show, however, a wide-range in plasma testosterone concentrations. These levels are not stable but are characterized by irregularly occurring fluctuations superimposed on a circadian rhythm with testosterone reaching peak concentrations during early morning (18). There is evidence that dysphoric states and psychosocial stress may depress plasma testosterone (19, 20). When these factors and time of sampling are controlled, a significant degree of intraindividual consistency in serum testosterone has been observed (21). At present, however, there is no evidence for a relation, within or across normal men, between circulating androgen levels and differences in sexual behavior (22). It deserves mention that most studies have included few subjects, have used an insufficient number of hormonal assessments, or have relied on brief questionnaires or psychological instruments of questionable reliability. In addition, most investigators have measured total (bound) testosterone and not the physiologically active free testosterone.

In aging men there is a tendency toward elevated serum luteinizing hormone (LH) and decreased bound and free plasma testosterone (23, 24). There are wide individual hormonal differences, however, and the covariation between circulating androgens and sex behavior during aging is presently not known.

Activation of receptor sites located in central and peripheral target tissues is necessary for the induction of physiological and behavioral changes. It is possible that maximum behavioral effects are obtained at relatively low concentration of circulating androgen and that beyond these levels significant hormonal behavioral relations are no longer evident. In addition, sexual functioning may not only reflect variations in plasma testosterone but also little-understood molecular mechanisms such as threshold changes in receptor site sensitivity.

ERECTILE DISORDERS

Any disease associated with general distress, pain, or fatigue or accompanied by fear, anxiety, or depression can potentially affect sexual function. These physical and emotional concomitants are frequently accompanied by diminished sexual desire, which may result in decreased sexual activity or in problems of sexual performance or satisfaction.

A number of disease states may also affect performance without directly interfering with sexual drive. The pathophysiology usually involves impairment of anatomical structures or physiological processes that mediate sexual function. Frequently, the sexual difficulties associated with medical illnesses are multidetermined and a full etiological understanding requires conceptual and methodological approaches from various frames of reference. We limit this discussion to neurovascular and hormonal factors that either singly or jointly impair erectile capacity.

Neurovascular Factors

Relatively little is known about the relation between specific brain lesions or abnormal brain activity and disturbances in sexual function. For over a century there have been reports that traumatic damage of the temporal lobe and temporal lobe epilepsy are associated with erectile impotence, hypoactive or hyperactive sexual desire, and deviant sexual behavior (25–28). Most of these reports are based on unsystematic observations that have not controlled for the role of cognitive defects, psychological or interpersonal difficulties, or the pharmacological effects of medication.

The systematic observations of patients with spinal cord injuries, as in the classic study of Bors & Comarr (12), have contributed significantly to our understanding of the neurophysiology of human sexual function. In an extensive investigation of 529 patients with spinal cord injuries, Bors & Comarr noted that when there is a complete lesion in the upper motor neuron, reflexogenic erections continue to occur since the pudendal sacral arc is not impaired. When there is a complete lower motor neuron lesion, reflex erections no longer take place but 20–30% of these patients retain the capacity to have psychologically induced erections via the thoracolumbar sympathetic outflow. A small proportion of them also ejaculate and achieve orgasmic release. When the lesions in the upper or lower motor neuron are incomplete, more than 90% of patients have erections; in these cases ejaculation seems to be more vulnerable than erection, particularly when the lesions reside in the upper motor neuron.

Neurological disorders that involve the spinal cord such as multiple sclerosis are also frequently associated with sexual symptoms (29, 30). We know very little about the pathogenesis of erectile failure in this and related conditions; care should be exercised in extrapolating from knowledge derived from patients with spinal cord injuries.

Degenerative or traumatic lesions of the nervous supply to genital structures frequently result in erectile problems. In diabetes mellitus, the prevalence of erectile impotence is high, ranging from 27–58% primarily as a function of age (31). Severity of the diabetes, duration of the disease, and the type of antidiabetic medication does not appear to be related to the occurrence of sexual problems. It is currently thought that a visceral neuropathy involving the autonomic outflow to the genitalia may be the primary pathogenic determinant. This is based on observations of (*a*) a higher frequency of peripheral and autonomic neuropathies in diabetics with erectile impotence than in nonimpotent diabetics (32), (*b*) the association between neuropathic bladder disorder and diabetic impotence (33), (*c*) a prolonged bulbocavernosus reflex response latency in impotent diabetics (34), and (*d*) histological abnormalities in autonomic fibers within genital structures of impotent diabetics but not in control samples (35).

Recent studies have not supported the notion that endocrine factors contribute to diabetic impotence (36–39). Instead, there is growing awareness that obstructive vascular pathology may play a significant role in the development of erectile disorders associated with diabetes (39).

Increasing attention has recently been paid to vascular insufficiency as a possible cause of erectile problems. Any pathological condition that interferes with the arterial supply to the corpora cavernosa can result in erectile impairment. This includes occlusive vascular diseases of the lower aorta and the internal iliac branches, congenital vascular anomalies, and traumatic lesions to the pelvis or genital blood vessels (40–42). In addition, pathological changes that increase the venous drainage from the corpora may also contribute to deficient or nonsustained erections. These abnormalities may include fistulas between the glans and the corpora or incompetent valves within the deep dorsal penile vein (9, 43).

More frequently, vasculogenic impotence appears to be due to degenerative vascular pathology associated with aging or unknown factors. Some investigators by means of systematic histological studies of the penile arterial bed from post-mortem specimens have shown age-related fibrotic changes in blood vessel walls and the narrowing or obstruction of the arterial lumen by thrombi and calcium deposits (44). The significance of these changes in causing erectile disorders remains to be determined.

The development of surgical approaches for the correction of vasculogenic impotence and procedures for penile prosthetic implantation emphasizes the importance of objective methods for assessing organic impairment in erectile function. One procedure that is being utilized with increasing frequency is the nocturnal penile tumescent method (NPT), which is based on the recordings of the cycle of erections associated with rapid eye movement or dreaming sleep. Considerable data have been presented (45, 46) suggesting that in psychologically impotent men NPT is generally preserved and clearly exceeds coital erectile capacity; in cases of organic impotence, NPT is impaired in keeping with the patient's waking performance. This procedure is of value in revealing organic impairment, but it does not provide information on the pathogenesis involved.

The measurement of penile blood pressure and penile blood flow by means of mercury strain gauge recordings, spectrography, and doppler techniques provides a more direct method of identifying the role of vascular obstruction (47–49). Methods for x-ray visualization of the genital vascular system include pelvic arteriography (50), phalloarteriography (42), and percutaneous cavernosography (9). These invasive approaches, when clinically indicated, can provide valuable information, but they remain difficult to implement and interpret.

The possible neuropathic contribution to erectile impairment may be evaluated indirectly by cystometrography because the autonomic pathways that control erection and micturition are the same (33). The electrophysiological testing of the bulbocavernosus reflex may also contribute information on the integrity of the reflex arc that subserves erection (51). Note, however, that each of these tests offers only indirect etiological evidence. Since a considerable overlap frequently exists between the observations derived from impotent and normal men, an abnormal recording does not necessarily prove that vascular or neural factors are responsible for the erectile incapacity. These procedures should be considered not in isolation but as ancillary tools in the context of extensive psychological and clinical information.

Hormonal Factors

The assessment of pituitary gonadal function in men with erectile disorders, presumably free from organic disease, has provided inconsistent findings. Some investigators observed that impotent men had lower mean urinary and plasma testosterone concentrations than comparison groups; other workers failed to show abnormalities in circulating testosterone (22). Similarly, the hormonal activation of the pituitary gonadal axis of men with erectile dysfunctions by means of LH releasing hormone (LHRH) or human chorionic gonadotrophin has been found normal by some investigators (52) and impaired by others (53). Methodological differences in number, frequency, and time of blood sampling and inadequate matching of control groups may account for the discrepancy in findings. It is conceivable that erectile disorders include heterogeneous conditions, but their possible identification necessitates the systematic gathering of detailed clinical and developmental histories and the use of reliable psychometric instruments usually lacking in previous psychoendocrine reports.

In hormonal studies of functional impotence, it is important not to ascribe etiological significance to low circulating testosterone, even if its existence is corroborated, because it could reflect the psychological stress induced by the sexual problem, the effect of depressive mood, or the lack of sexual activity.

Despite a long history of clinical trials, there is still no solid evidence that androgen administration is effective for the treatment of "psychological" erectile dysfunctions (22). Drugs like clomiphene, a nonsteroid triethylene derivative that markedly elevates plasma testosterone, or synthetic androgens such as mesterolone have also not been found effective in the treatment of functional impotence. Some investigators have evaluated the effect of composite drugs that combine androgens with thyroid hormones or agents such as yohimbine or strychnine (22). Positive effects in the treatment of erectile disorders have been reported, but the interpretation of these studies is obscured by the lack of double-blind control protocols, the limited behavioral information provided, and the inclusion of aging men and hypogonadal patients as subjects. Bancroft (54) recently raised the possibility that exogenous androgens may interact positively with psychological modalities of treatment in the resolution of erectile symptoms. Evidence of such possible interaction has been observed in the treatment of female orgastic problems, but it remains to be demonstrated in men.

Recently, some investigators have begun experimenting with the use of LHRH in the treatment of impotent men; the results are equivocal. Mortimer et al (55), in a study of males with hypogonadism of hypothalamic or pituitary origin, found that administration of gonadotrophic releasing hormones resulted in a rise in circulating androgen levels and in a rapid increase in potency that occurred while plasma androgens were still well below the normal range. The authors speculated that LHRH may influence sexual behavior by a central action independent from hormonal effects.

Benkert et al (56) conducted a double-blind study comparing the effects of LHRH against placebo in a group of impotent men free from clinical evidence of endocrine pathology. A significantly better therapeutic response was observed in response to LHRH than placebo but only during a 4–6 week period after termination of the active medication. Davies et al (57), on the other hand, were unable to detect significant effects of LHRH on sexual activity in a double-blind crossover study of men with erectile impotence of undetermined cause. Clearly, additional studies are required before any conclusions can be drawn on the effect of LHRH on sexual function.

Diminished sexual desire and erectile capacity and ejaculatory problems are frequently noted in men suffering from various hypogonadal disorders. Male hypogonadism may be primary as a direct result of gonadal impairment or secondary to abnormalities in hypothalamic-pituitary activity. In instances of primary hypogonadism, impairment of testicular Leydig cells leads to low testosterone production and a compensatory increase of LH (hypergonadotropic hypogonadism). When the impairment occurs at the hypothalamic-pituitary level, subnormal secretion of LH and testosterone are usually observed (hypogonadotropic hypogonadism).

Several recent reports show an association between pituitary tumors, high serum prolactin concentrations, depressed testosterone secretion, and erectile impotence (58, 59). It is believed that the hypogonadism observed in men with hyperprolactinemia is secondary to a depression in gonadotrophic secretion. Administration of bromocryptine, a dopamine agonist, frequently lowers prolactin concentration, elevates LH and testosterone secretion, and restores sexual potency (60). The mechanism of action of bromocryptine in the treatment of erectile disorders associated with hyperprolactinemia is not fully understood. It may have a central effect on dopaminergic mechanisms that mediate sexual function or it may improve erectile capacity due to the increase in testosterone secretion that occurs as prolactin levels decrease. It has been noted, however, that restoration of erectile potency may occur before the normalization of circulating testosterone (61). Bromocryptine does not seem effective in the treatment of erectile disorders unless they are associated with hyperprolactinemia (62).

A recent study of 105 consecutive patients complaining of erectile difficulties who were screened for serum testosterone revealed that 35% of them had unsuspected disorders of the hypothalamic-pituitary-gonadal axis (63). Twenty patients had hypogonadotropic hypogonadism mainly due to pituitary tumors, seven had hypergonadotropic hypogonadism due to primary gonadal failure, and eight suffered from hyperprolactinemia. Although the prevalence of hormonal dysfunctions seems surprisingly high and may reflect a nonrandom sample, this investigation emphasizes the importance of measuring plasma testosterone during the routine screening of patients with erectile disorders.

The behavioral consequences of hypogonadism are quite varied and in some men sexual desire and the capacity for sexual performance may remain unchanged for years. Clinical experience suggests that factors such as age at onset of hypogonadism, level of premorbid sexual competency, the medical and psychiatric status of the subject, and the attitude of the sexual partner are of considerable importance in determining the extent of the sexual impairment. At present few studies have systematically assessed clinical, hormonal, and psychological factors in relation to various sexual behavior dimensions in hypogonadal men.

In contradistinction with functionally impotent patients, administration of androgen to men with hypogonadal disorders frequently restores sexual desire and potency (64, 65). The literature in this area, however, is limited to case reports or to group studies that do not control for experimental bias or placebo effects. Recently, Davidson et al (66) reported the first double-blind investigation on the effects of a long-acting testosterone preparation in the treatment of six adult hypogonadal men. Significant dose-behavior relations were demonstrated between frequency of spontaneous and nocturnal erections, frequency of coitus, and androgen administration. There is considerable need for studies of this kind.

CONCLUSION

The current interest in the evaluation and treatment of erectile disorders brings into focus our limited knowledge about the physiology of male sexual function. There is growing awareness derived from several lines of physiological evidence that male erectile disorders include a wide range of heterogeneous conditions. The recent development of ancillary diagnostic methods underlines the importance of neurovascular and endocrine factors in the pathophysiology of erectile disorders associated with medical illnesses. The development and rational application of treatment methods should rest on conceptual and methodological approaches that integrate psychological and biological perspectives.

ACKNOWLEDGMENT

Supported by research grants MH 27513 and AM 20845, US Public Health Service. I thank Ms. Natalie Weinstein for her assistance in the preparation of this manuscript.

Literature Cited

1. Kinsey, A. C., Pomeroy, W. B., Martin, C. F. 1948. *Sexual Behavior in the Human Male.* Philadelphia; Saunders. 804 pp.
2. Weiss, H. D. 1972. The physiology of human penile erection. *Ann. Intern. Med.* 75:793–99
3. deGroat, W. C., Booth, A. M. 1980. Physiology of male sexual function. *Ann. Intern. Med.* 92 (Part 2):329–31
4. MacLean, P. D. 1966. Studies on the cerebral representation of certain basic sexual functions. In *Brain and Behavior,* ed. R. A. Gorski, R. E. Whalen, 3:35–79. Los Angeles: Univ. Calif. Press. 289 pp.
5. Heath, R. G. 1972. Pleasure and brain activity in man. *J. Nerv. Ment. Dis.* 154:3–17
6. Gessa, G. L., Tagliomonte, A. 1974. Possible role of brain serotonin and dopamine in controlling male sexual behavior. *Adv. Biochem. Psychopharmacol.* 2:217–28
7. Whalen, R. E., Gorzalka, B. B., DeBold, J. F. 1975. Methodological considerations in the study of animal sexual behavior. In *Sexual Behavior: Pharmacology and Biochemistry,* ed. M. Sandler, G. L. Gessa, pp. 33–44. New York: Raven. 354 pp.
8. Conti, G. 1952. L'érection du pénis humain et ses bases morphologicovasculaires. *Acta Anat.* 14:217–62
9. Fitzpatrick, T., Cooper, J. F. 1975. A cavernosogram study on the vascular competence of the human deep dorsal vein. *J. Urol.* 113:497–99
10. O'Conor, V. J. 1958. Impotence and the Leriche syndrome: an early diagnostic sign; consideration of the mechanism; relief by endarterectomy. *J. Urol.* 80:195–97
11. Shirai, M., Ishii, N., Mitsukaw, S., Matsuda, S., Nakamura, M. 1978. Hemodynamic mechanism of erection in the human penis. *Arch. Androl.* 1:345–49
12. Bors, E., Comarr, A. E. 1960. Neurological disturbances of sexual function with special reference to 529 patients with spinal cord injury. *Urol. Surv.* 10:191–222
13. Wagner, G., Brindley, G. S. 1978. *The effect of atropine, alpha and beta blockers upon penile erection. A controlled pilot study.* Presented at 1st Int. Conf. Vasculogenic Impotence, Sept., 1978, New York
14. Baumgarten, H. G., Falck, B., Lange, W. 1969. Adrenergic nerves in the corpora cavernosa penis of some mammals. *Z. Zellforsch.* 95:58–67
15. Melman, A., Henry, M. D. 1979. The possible role of the cathecholamines of the corporal bodies in penile erection. *J. Urol.* 121:419–21
16. McConnell, J., Benson, G. S., Wood, J. 1979. Autonomic innervation of the mammalian penis: a histochemical and physiological study. *J. Neural Transm.* 45:227–38
17. Domer, F. R., Wessler, G., Brown, R. L., Charles, H. C. 1978. Involvement of the sympathetic nervous system in the urinary bladder, internal sphincter and in penile erection in the anesthetized cat. *Invest. Urol.* 15:404–7
18. Schiavi, R. C., Davis, D. M., White, D., Edwards, A., Igel, G., Fisher, C. 1974. Plasma testosterone during nocturnal sleep in normal men. *Steroids* 24:191–201
19. Doering, C. H., Brodie, H. K. H., Kraemer, H. C., Moos, R. H., Becker, H. B., Hamburg, D. 1975. Negative affect and plasma testosterone: a longitudinal human study. *Psychosomatic Medicine* 37:484–91
20. Kreuz, L. E., Rose, R. M., Jennings, J. R. 1972. Suppression of plasma testosterone levels and psychological stress. *Arch. Gen. Psychiatry* 26:479–83
21. Rose, R. M., Kreuz, L. E., Holaday, J. W., Sulak, K. J., Johnson, C. E. 1972. Diurnal variation of plasma testosterone and cortisol. *J. Endocrinol.* 54:177–78
22. Schiavi, R. C., White, D. 1976. Androgens and male sexual function: A review of human studies. *J. Sex Marital Ther.* 2:214–28
23. Isurugi, K., Fukutani, K., Takayasu, H., Wakabayashi, K., Tamaoki, B. 1974. Age-related changes in serum luteinizing hormone (LH) and follicle-stimulating hormone (FSH) levels in normal men. *J. Clin. Endocrinol. Metab.* 39:955–57
24. Stearns, E. L., MacDonnell, J. A., Kaufman, B. J., Padva, R., Lucman, T. S., Winter, J. S. D., Faiman, C. 1974. Declining testicular function with age: hormonal and clinical correlates. *Am. J. Med.* 57:761–66
25. Taylor, D. C. 1969. Sexual behavior and temporal lobe epilepsy. *Arch. Neurol.* 21:510–16
26. Saunders, M., Rawson, M. 1970. Sexuality in male epileptics. *J. Neurol. Sci.* 10:577–83

27. Weinstein, E. A. 1974. Sexual disturbances after brain injury. *Med. Aspects Hum. Sex.* 8:10–31
28. Blumer, D. 1970. Hypersexual episodes in temporal lobe epilepsy. *Am. J. Psychiatry* 126:1099–1106
29. Horenstein, S. 1976. Sexual dysfunction in neurological disease. *Med. Aspects Hum. Sex.* 10:6–13
30. Vas, C. J. 1969. Sexual impotence and some autonomic disturbances in men with multiple sclerosis. *Acta Neurol. Scand.* 45:166–82
31. Schiavi, R. C., Hogan, B. 1979. Sexual problems in diabetic mellitus: psychological aspects. *Diabetes Care* 2:9–17
32. Campbell, I. W. 1976. Diabetic autonomic neuropathy. *Br. J. Clin. Pract.* 30:153–56
33. Ellenberg, M. 1971. Impotence in diabetes: the neurologic factor. *Ann Intern. Med.* 75:213–19
34. Vacek, J., Lachman, M. 1977. Bulbokavernózni reflex u diabetikůs poruchou erectivity: Klinická e elektromyografická studie. *Cas. Lek. Cesk.* 33:1014–17
35. Faerman, I., Glacer, L., Fox, D., Jadzinsky, M. N., Rappaport, M. 1972. Impotence and diabetes: histological studies of the autonomic nervous fibers of the corpora cavernosa in impotent diabetic males. *Diabetes* 23:971–76
36. Kolodny, R. C., Kahn, C. B., Goldstein, H. H., Barnett, D. M. 1973. Sexual dysfunction in diabetic men. *Diabetes* 23:306–9
37. Rastogi, G. K., Chakraborti, J., Sinha, M. K. 1974. Serum gonadotropins (LH and FSH) and their response to synthetic LHRH in diabetic men with and without impotence. *Horm. Metab. Res.* 6:335–36
38. Wright, A. D., London, D. R., Holder, G., Williams, J. W. Rudd, B. T. 1976. Luteinizing release hormone tests in impotent diabetic males. *Diabetes* 25: 975–78
39. Herman, A., Adar, R., Rubinstein, Z., Hashomer, T., Aviv, T. 1978. Vascular lesions associated with impotence in diabetic and non-diabetic arterial occlusive disease. *Diabetes* 27:975–81
40. Leriche, R., Morel, A. 1948. The syndrome of thrombotic obliteration of the aortic bifurcation. *Ann. Surg.* 127:193–206
41. Michal, V., Kramar, R., Bartak, V. 1974. Femoro-pudendal bypass in the treatment of sexual impotence. *J. Cardiovasc. Surg.* 15:356–59
42. Michal, V., Pospíchal, J. 1978. Phalloarteriography in the diagnosis of erectile impotence. *World J. Surg.* 2:239–47
43. Ebbehøj, J., Wagner, G. 1979. Insufficient penile erection due to abnormal drainage of the cavernous bodies. *Urology* 13:507–10
44. Ružbarský, V., Michal, V. 1977. Morphologic changes in the arterial bed of the penis with aging. *Invest. Urol.* 15:194–99
45. Karacan, I., Salis, P. J., Williams, R. L. 1978. The role of the sleep laboratory in diagnosis and treatment of impotence. In *Sleep Disorders: Diagnosis and Treatment,* ed. R. L. Williams, I. Karacan, pp. 353–82. New York: Wiley. 452 pp.
46. Fisher, C., Schiavi, R. C., Edwards, A., Davis, D. M., Reitman, M., Fine, J. 1979. Evaluation of nocturnal penile tumescence (NPT) in the differential diagnosis of sexual impotence: a quantitative study. *Arch. Gen. Psychiatry* 36:431–37
47. Gaskell, P. 1971. The importance of penile blood pressure in cases of impotence. *Can. Med. Assoc. J.* 105:1047–51
48. Abelson, D. 1975. Diagnostic value of penile pulse and blood pressure: Doppler study of impotence in diabetics. *J. Urol.* 113:636–39
49. Engel, G., Burnham, S. J., Carter, M. F. 1978. Penile blood pressure in the evaluation of erectile impotence. *Fertil. Steril.* 30:687–90
50. Michal, V., Kramář, R., Popischal, J., Hejhal, L. 1977. Arterial epigastricocavernous anastomosis for the treatment of sexual impotence. *World J. Surg.* 1:515–20
51. Ertekin, C., Reel, F. 1976. Bulbocavernosus reflex in normal men and in patients with neurogenic bladder and/or impotence. *J. Neurol. Sci.* 28:1–12
52. Ismail, A. A. A., Davidson, D. W., Loraine, J. A. 1970. Assessment of gonadal function in impotent men. In *Reproductive Endocrinology,* ed. W. J. Irvine, pp. 138–47. Edinburgh: Churchill Livingstone. 385 pp.
53. Delitalia, G., Masala, A., Alagna, S., Lotti, G. 1977. Luteinizing hormone, follicle stimulating hormone and testosterone in normal and impotent men following LH-RH and HCG stimulation. *Clin. Endocrinol.* 6:11–15
54. Bancroft, John. 1977. Editorial. Hormones and sexual behavior. *Psychol. Med.* 7:553–56
55. Mortimer, C. H., McNeilly, A. S., Fisher, R. A., Murray, M. A. F., Besser, G. M. 1974. Gonadotrophin releasing

hormone therapy in hypogonadal males with hypothalamic or pituitary dysfunction. *Br. Med. J.* 4:617–21

56. Benkert, O., Jordan, R., Dahlen, H. G., Schneider, H. P. G., Gammel, G. 1975. Sexual impotence: a double-blind study of LH-RH nasal spray versus placebo. *Neuropsychobiology* 1:203–10
57. Davies, T. F., Mountjoy, C. Q., Gomez-Pan, A., Watson, M. J., Hanker, J. P., Besser, G. M., Hall, R. 1976. A double-blind crossover trial of LHRH in sexually impotent men. *Clin. Endocrinol.* 5:601–7
58. Carter, J. N., Tyson, J. E., Tolis, G., Van Vliet, S., Faiman, C., Friesen, H. G. 1978. Prolactin-secreting tumors and hypogonadism in 22 men. *N. Engl. J. Med.* 299:847–52
59. Snyder, P. J., Bigdeli, H., Gardner, D. F., Mihailovic, V., Rudenstein, R. S., Sterling, F. H., Utiger, R. D. 1979. Gonadal function in fifty men with untreated pituitary adenomas *J. Clin. Endocrinol. Metab.* 48:309–14
60. Ambrosi, B., Bara, R., Travaglini, P., Weber, G., Peccoz, P. B., Rondena, M., Elli, R., Faglia, G. 1977. Study of the effects of bromocriptine on sexual impotence. *Clin. Endocrinol.* 7:417–21
61. McGregor, A. M., Scanlon, M. F., Hall, K., Cook, D. B., Hall, R. 1979. Reduction in size of a pituitary tumor by bromocriptine therapy. *N. Engl. J. Med.* 300:291–93
62. March, C. M. 1979. Bromocriptine in the treatment of hypogonadism and male impotence. *Drugs* 17:349–58
63. Spark, R. F., White, R. A., Connolly, P. E. 1980. Impotence is not always psychogenic. Newer insights into hypothalamic-pituitary gonadal dysfunction. *J. Am. Med. Assoc.* 243:750–55
64. Huffer, V., Scott, W. M., Conner, T. B., Lovice, H. 1964. Psychological studies of adult male patients with sexual infantilism before and after androgen therapy. *Ann. Intern. Med.* 61:255–68
65. Beaumont, P. J. V., Bancroft, J. H. J., Beardwood, C. J., Russell, G. F. M. 1972. Behavioral changes after treatment with testosterone: case report. *Psychol. Med.* 2:70–72
66. Davidson, J. M., Camargo, C. A., Smith, E. R. 1979. Effects of androgen on sexual behavior in hypogonadal men. *J. Clin. Endocrinol. Metab.* 48:955–58

Ann. Rev. Med. 1981. 32:521–42
Copyright © 1981 by Annual Reviews Inc. All rights reserved

DISORDERS OF GLUTAMATE METABOLISM AND NEUROLOGICAL DYSFUNCTION

❖7417

Stanley B. Prusiner, M.D.

Howard Hughes Medical Institute Laboratory, Departments of Neurology
and Biochemistry and Biophysics, University of California, San Francisco,
California 94143

Introduction

The metabolism of glutamate in man, animals, and bacteria has been widely investigated in recent years. From these studies an impressive array of disorders of glutamate metabolism has emerged (Table 1). Most of the disorders appear to be associated with or even responsible for dysfunction of the central nervous system (CNS) in man. This is not surprising since glutamate and its metabolites play an important role in the intermediary metabolism of the brain, as well as in chemical neurotransmission at synapses (1–9). Although all of the disorders discussed here have been recorded in literature, it has not been widely appreciated that this whole family of disorders is frequently associated with neurological dysfunction.

Pathways of Glutamate Metabolism

Three metabolites, α-ketoglutarate, glutamate, and glutamine, form not only the center of cellular nitrogen metabolism, but also a crossroads where carbohydrate and nitrogen metabolism intersect (10). In Figure 1 the structures of these metabolites are illustrated with the enzymes that catalyze their interconversion. Each of the three metabolites is composed of five carbon atoms and each contains a carboxyl group adjacent to the α-carbon.

α-KETOGLUTARATE α-Ketoglutarate has two major metabolic fates. First is the decarboxylation of α-ketoglutarate, which is catalyzed by the α-ketoglutarate dehydrogenase complex, resulting in the formation of succinyl CoA. The catalytic activity of the α-ketoglutarate dehydroge-

0066-4219/81/0401-0521$01.00

Table 1 Disorders of glutamate metabolism[a]

1. Pyroglutamic aciduria and glutathione synthetase deficiency
2. Glutathionuria and γ-GTP deficiency
3. Glutathione reductase deficiency
4. Glutathione peroxidase deficiency
5. γ-Glutamyl cysteine synthetase deficiency
6. Cysteinyl glycinuria
7. Type I hyperprolinemia and proline oxidase deficiency
8. Type II hyperprolinemia
9. Histidinemia
10. Homocarnosine accumulation
11. Vitamine B_6-dependent seizures and glutamate decarboxylase
12. Huntington's chorea and glutamate decarboxylase
13. Hyper-β-alaninemia and GABA aciduria
14. Hereditary protein intolerance and glutaminase
15. Carbamyl-phosphate synthetase deficiency
16. Hepatic insufficiency and encephalopathy
17. Glutamate intolerance and toxicity

[a] All of these disorders are associated with known neurological dysfunction except glutathione peroxidase deficiency.

nase complex is subject to rigorous control by phosphorylation and metabolite effectors (11). The second fate of α-ketoglutarate involves the formation of glutamate upon addition of an ammonium ion group to the α-carbon of α-ketoglutarate; this reaction is catalyzed by the enzyme glutamate dehydrogenase. The ready reversibility of this amidation reaction has been extensively studied and appears to provide a mechanism for modulating the flow of α-ketoglutarate into either carbohydrate metabolism via the Krebs cycle or into nitrogen metabolism via the glutamate dehydrogenase pathway (12, 13). The levels of α-ketoglutarate are also controlled by a transaminase reaction, whereby the α-amino group of an amino acid is transferred to the α-carbon of α-ketoglutarate to form glutamate.

GLUTAMATE The metabolism of glutamate is considerably more diverse than that of α-ketoglutarate. The carbon skeleton of glutamate can readily be converted into a wide variety of different metabolites, including α-ketoglutarate, glutamine, pyroglutamate, γ-aminobutyric acid (GABA), α-glutamyl peptides, and γ-glutamyl peptides. The conversion of glutamate to glutamine is of major importance in cellular metabolism. The process requires one mole of ATP for each mole of ammonium ion incorporated during the formation of glutamine. The reaction is catalyzed by glutamine synthetase and, unlike the interconversion between α-ketoglutarate and glutamate, this reaction is not readily reversible. A separate enzyme, glutaminase, catalyzes the deamidation of glutamine to form glutamate and ammonium ions in the presence of water. In most organisms, glutamate is the most abundant intracellular amino acid. This is true for single-cell bacteria as well as complex mammalian systems with multiple organs (14).

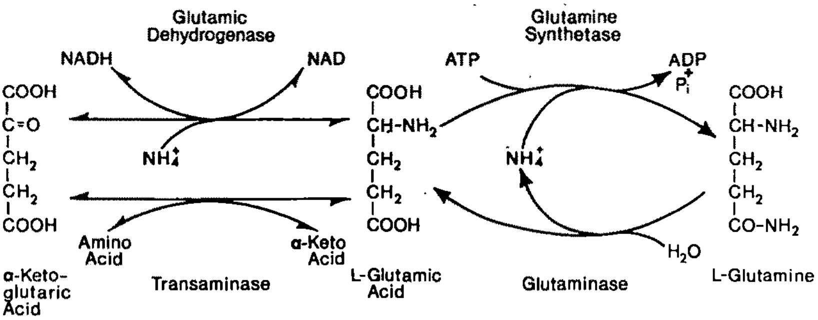


Figure 1 Core reactions of glutamate metabolism showing enzyme-catalyzed interconversions of α-ketoglutarate, glutamate, and glutamine.

Glutamate exists at rather low concentrations extracellularly whereas glutamine is one of the most abundant extracellular amino acids in most mammals. As in most cells, high concentrations of glutamic acids are found intracellularly in the mammalian brain while high concentrations of glutamine can be found both in the blood and the cerebrospinal fluid (CSF) (Table 2).

Since the conversion of glutamate to glutamine requires ATP, the synthesis and degradation of glutamine must be rigorously controlled in order to prevent the coupling of glutamine synthetase and glutaminase reactions from forming a "futile cycle" of amide synthesis and degradation (15) (Figure 1). If such a cycle were allowed to proceed uncontrolled, wasteful consumption of cellular energy stores would result. Relatively little is known about the control mechanisms that modulate the catalytic activities of these two enzymes in the nervous systems of animals, but extensive information exists in bacterial systems, so these data are briefly reviewed below.

An extraordinary degree of modulation is imposed upon the glutamine synthetase reaction in *Escherichia coli* by a series of enzymatically catalyzed reactions that involve the covalent modification of the enzyme whereby each of the 12 subunits of the enzyme can potentially become adenylylated (16, 17). The adenylylation reaction is performed by an enzyme complex, the activity of which is modulated by uridylylation of one component of the enzyme complex and by the concentrations of α-ketoglutarate and glutamine (16, 17). The adenylylation of glutamine synthetase (*a*) directly modulates the catalytic activity of the enzyme, (*b*) increases the sensitivity of the enzyme to feedback inhibition by biosynthetic end products, and (*c*) changes the affinity of the enzyme for divalent cations (16, 17). Attempts to demonstrate covalent modification of glutamine synthetase in mammals

Table 2 Concentrations of glutamate metabolites in the human nervous system

Metabolites	Brain[b] (μmoles/g)	Plasma[c] (μM)	CSF[d] (μM)
Glutamate	7.1	22	1.8
Glutamine	5.6	610	600
Ammonia		40	18
GABA	0.59	<0.02	<0.02
Glutathione[a]	2.1	<0.5	<0.5
α-Ketoglutaramate			6.0
Pyroglutamate			17

[a] Represents total glutathione calculated as GSH (158, 159).
[b] Values for frontotemporal cortex biopsies frozen in liquid nitrogen (159).
[c] Data from (158, 160).
[d] Data from (158, 160, 161).

have failed to date (18, 19). In some bacteria, the levels of glutamine synthetase are regulated by the availability of nitrogenous metabolites and by cAMP (20, 21). Recent investigations have begun to unravel the complex genetic regulation of glutamine synthetase (22, 23). Glutamine synthetase, besides regulating its own synthesis, also appears to regulate the synthesis of several enzymes involved in the formation of glutamate (24–26) and even possibly to regulate sporulation (27). The synthesis of glutamate dehydrogenase is diminished by glutamine synthetase, while the synthesis of enzymes involved in histidine and proline catabolism is enhanced by both glutamine synthetase and cAMP. Glutamine synthetase also increases the synthesis of a periplasmic asparaginase, which exhibits γ-glutamyl transpeptidase activity (28). The identification of ColE1 hybrid plasmids containing *E. coli* genes coding for glutamine synthetase, glutamate dehydrogenase, and glutamate synthase should provide interesting new tools for studying the regulatory mechanisms controlling nitrogen assimilation (29).

Glutaminase B from *E. coli,* which directly opposes the action of glutamine synthetase, is modulated by carboxylic acids and adenine nucleotides (15, 30). The enzyme is inhibited by ATP and ADP, but activated by AMP. In contrast, the glutamine synthetase reaction utilizes ATP as a substrate, while AMP inhibits the reaction. A system of reciprocal control appears to regulate the activities of these two enzymes and thus prevents the occurrence of a futile cycle of amide synthesis and degradation (15).

It would not be surprising if the enzymatic regulation of glutamate and glutamine synthesis and degradation in the mammalian CNS were just as sophisticated and complex as that found in many bacteria, especially in view of the high concentrations of glutamate metabolites in the nervous system (Table 2). Although an enormous amount of literature on the metabolism of glutamate in the CNS exists, relatively few studies have focused on the metabolic mechanisms that regulate the levels of glutamate metabolites. Studies with radioisotopically labelled compounds have shown that multiple, distinct pools of glutamate and its metabolites exist in brain (1, 31). The localization of these pools remains unclear: they may reflect different cell types, such as neurons, astroglia, and oligodendroglia, and/or possibly different subcellular compartments (1, 31, 32). Considerable evidence suggests that at least one small subcellular compartment of glutamate may be involved in neurotransmission. Glutamate is thought to be an excitatory neurotransmitter in the mammalian CNS, but firm evidence in support of this notion is lacking. However, several studies in invertebrates have provided good evidence for the role of glutamate as an excitatory neurotransmitter (33, 34). In support of the putative role of glutamate in neurotransmission of the mammalian CNS are studies showing (*a*) the Ca^{2+}-dependent release of glutamate from synaptosomal fractions after electrical stimulation or elevation of the potassium concentration of the incubation media, (*b*) the high affinity binding of glutamate to subcellular fractions of brain, and (*c*) the application of glutamate to neural membranes resulting in their depolarization (34–36). Studies in support of glutamate as a neurotransmitter are hampered by the high concentrations of glutamate in nervous system cells, most of which is almost certainly *not* involved in neurotransmission. Until techniques become available to isolate neurons and their synaptic terminals, in which glutamate functions as a neurotransmitter, the role of glutamate in neurotransmission will continue to be putative (37, 38).

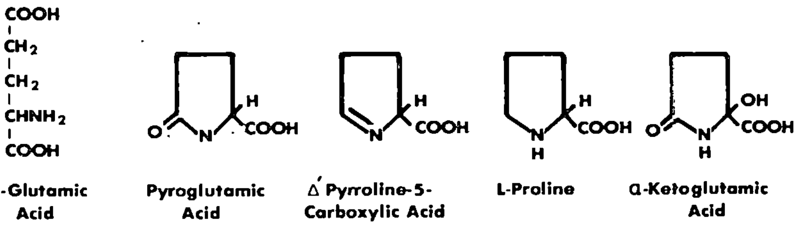
GABA Another pathway of glutamate metabolism in the nervous system involves the decarboxylation of glutamate to form GABA. In contrast to glutamate, which may function as an excitory neurotransmitter, GABA functions as an inhibitory neurotransmitter (39), as evidenced by studies on the nervous system of invertebrates (33–35, 39). A similar role for GABA in the mammalian CNS is now reasonably well established (9, 39, 40) by studies describing (*a*) the nonuniform distribution of GABA in the CNS where high concentrations are localized mainly in synaptosomal fractions from specific areas of the brain, (*b*) the high affinity binding of GABA to preparations enriched for synaptic terminals, (*c*) the Ca^{2+}-dependent release of GABA from synaptosomes by electrical stimulation, (*d*) the hyperpolarization of neurons exposed to GABA antagonized by picrotoxin and bicuculline, and (*e*) the benzodiazepine receptor interacting with GABA receptor and enhancing the inhibitory effect of GABA (34, 39–43). The endogenous ligand for the benzodiazepine receptor is unknown but inosine as well as nicotinamide are candidates (43, 44). Nervous mutant mice, who have a selective loss of more than 90% of their cerebellar Purkinje cells during maturation, show a marked reduction in the number of cerebellar benzodiazepine binding sites (45).

GABA is degraded to succinate in a sequence of two reactions. First, succinic semialdehyde is formed when the α-amino moiety of GABA is transferred to α-ketoglutarate in a transamination reaction. Reduction of succinic semialdehyde leads to the formation of γ-hydroxybutyric acid; oxidation of the aldehyde leads to the formation of succinate in a NAD-dependent reaction. High concentrations of GABA are found in the nervous systems of both invertebrates and vertebrates (Table 2), but they are not unique to nervous tissue since large amounts have also been found in the mammalian kidney. The function of GABA in kidney remains elusive. High concentrations of glutamate and glutamine are also found in kidney tissue, where they are involved in the excretion of ammonium ions.

PYROGLUTAMATE AND GLUTATHIONE Cyclization of glutamate results in the formation of pyroglutamate (Figure 2). The formation of pyroglutamate or 5-oxoproline occurs via at least four enzyme-catalyzed reactions: (*a*) γ-glutamyl cyclotransferase, (*b*) a side reaction of glutamine synthetase in the absence of ammonia, (*c*) a partial reaction of γ-glutamyl cysteine synthetase in the absence of cysteine, and (*d*) formation of N-terminal pyroglutamate in proteins (46). Recent studies show that pyroglutamate is converted in an ATP-dependent reaction catalyzed by the enzyme 5-oxoprolinase (47).

In addition to pyroglutamic acid, three other cyclized derivatives of glutamic acid are shown in Figure 2. Δ^1-Pyrroline-5-carboxylic acid is an intermediate in the synthesis and degradation of the amino acid proline. α-Ketoglutaramate is formed in a transamination reaction, where the alpha amino group of glutamine is transferred to an α-keto acid such as α-ketoglutarate (see below).

A major route of pyroglutamate formation appears to be the cleavage of γ-glutamyl peptides by γ-glutamyl cyclotransferase (48). A variety of natu-



L-Glutamic Acid

Pyroglutamic Acid

Δ^1 Pyrroline-5-Carboxylic Acid

L-Proline

α-Ketoglutamic Acid

Figure 2 Cyclized derivatives of glutamic acid.

rally occurring γ-glutamyl peptides have been described; presumably their synthesis is catalyzed by the enzyme γ-glutamyl transpeptidase (49, 50). This enzyme can transfer a γ-glutamyl moiety from a variety of donor γ-glutamyl peptides to an assortment of acceptor oligopeptides and amino acids (51). The physiological functions of the enzyme γ-glutamyl transpeptidase remain unclear. Several investigators suggest that the enzyme primarily functions in the degradation of glutathione (52–57), while others think it translocates peptides and amino acids across cell surfaces (58–63). The possibility that translocation of amino acids might be coupled to the γ-glutamyl cycle in which glutathione degradation and resynthesis occurs has aroused considerable interest (5, 6, 60, 64). The scheme requires three moles of ATP be consumed for each mole of peptide or amino acid transported. An alternative hypothesis suggests that group translocation occurs with the repetitive transfer of γ-glutamyl moieties, where the transported γ-glutamyl acceptor product subsequently becomes a γ-glutamyl donor substrate (63). Since reutilization of the γ-glutamyl product requires relatively little energy, such a system would be considerably more efficient than the γ-glutamyl cycle. Whether or not γ-glutamyl transpeptidase actually participates in the translocation of oligopeptides, and possibly some amino acids, remains to be established. Inhibition of γ-glutamyl transpeptidase in mice caused increased excretion of glutathione but no amino aciduria was observed (65). Similar observations have been made in humans suffering from γ-glutamyl transpeptidase deficiency as described below. These studies certainly favor a degradative role for the enzyme. The high degree of specificity that γ-glutamyl transpeptidase from choroid plexus exhibits toward oligopeptide acceptors suggests that the enzyme may function in controlling the availability of biologically active peptides (66). After nearly a decade of intensive investigation, the many puzzles still surrounding γ-glutamyl transpeptidase testify to the need for its continued investigation (6, 53, 67).

The cellular regulation of γ-glutamyl transpeptidase in rat mammary tissue is controlled by prolactin (68). γ-Glutamyl transpeptidase in isolated brain endothelial cells and choroid plexus epithelial cells decreases dramatically as a function of time in culture (69, 70). Restoration of the enzyme in endothelial cells was induced by coculture of glial cells (69). Recent investigations on isolated kidney cells have suggested that γ-glutamyl transpeptidase may be intimately involved in the oxidation of reduced glutathione (71). At present, it is unclear whether this oxidative process is a consequence of γ-glutamyl transpeptidase activity or whether the oxidation is catalyzed by a separate enzyme (72, 73).

Glutathione is the most abundant natural γ-glutamyl peptide; its synthesis occurs in two sequential reactions catalyzed by the enzymes γ-glutamyl cysteine synthetase and glutathione synthetase (64, 74). Each synthetase-catalyzed reaction consumes one mole of ATP for each mole of product formed. Of interest is the isolation of mutants of *E. coli* containing no detectable reduced glutathione (GSH) and exhibiting a deficiency of either glutathione synthetase or γ-glutamyl cysteine synthetase (75). The growth rates in minimal medium without GSH supplementation were normal.

GLUTAMINE The amide nitrogen of glutamine has many metabolic fates in the biosynthesis of amino sugars, nucleotides, and amino acids (10). Removal of the amide nitrogen of glutamine results in the formation of glutamate, while removal of the α-amino nitrogen in a transamidation reaction yields α-ketoglutaramate. The formation of α-ketoglutaramate is catalyzed by the enzyme glutamine transaminase, but the function of α-ketoglutaramate in intermediary metabolism remains unclear (76). α-Ketoglutaramate appears to exist predominantly as a lactam ring in equilibrium with an open-chain form that undergoes deamidation in a reaction catalyzed by ω-amidase (77). In addition to the wide range of metabolic reactions described above, in which glutamine participates, the condensation of two molecules of glutamine to form γ-glutamyl glutamine is catalyzed by γ-glutamyl transpeptidase. Also of interest are the transglutaminases, which catalyze the formation of ϵ(γ-glutamyl) lysine cross-links in the polymerization of fibrin during blood clotting and the formation of vaginal plugs (78). Transglutaminases appear to be essential in receptor-mediated endocytosis of α_2-macroglobulin and some polypeptide hormones; they also catalyze the incorporation of polyamines into proteins through a γ-glutamyl linkage (79, 80).

Disorders of Glutamate Metabolism

A detailed map of glutamate metabolism with its associated disorders is shown in Figure 3. The numbers in this map designate the location of enzymatic defects or metabolite accumulations occurring in these disorders, and correspond to those listed in Table 1. Although a rigorous scheme of metabolic control undoubtedly exists, only a few components of the regulatory system are now known. A complete understanding of this complex area of metabolism in the nervous system, which is so central to all cellular nitrogen and carbohydrate metabolism, will undoubtedly require a much greater body of knowledge than is currently available. Perhaps the most striking aspect of this metabolic map is the fact that many separate, distinguishable disorders of glutamate metabolism are associated with dysfunction of the CNS. For most of the disorders described below, only a few patients have been identified and studied. A critical evaluation of the data must await further observations, which will either confirm or alter our

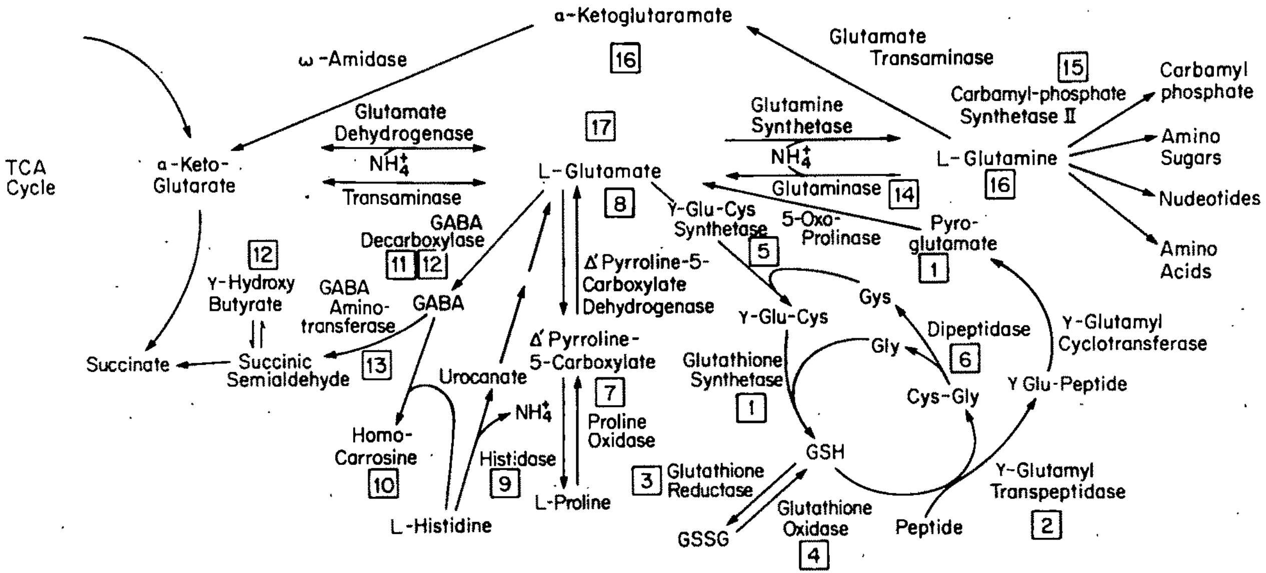


Figure 3 Pathways of glutamate metabolism and its associated disorders. Numbers inside squares refer to the disorders listed in Table 1, signifying abnormal levels of metabolites or altered enzyme function.

understanding of these diseases. It is hoped that this catalogue of disorders of glutamate metabolism will stimulate the identification and study of new cases and possibly new diseases at this important crossroads of cellular carbohydrate and nitrogen metabolism.

The finding of an abnormal accumulation of a metabolite does not necessarily imply a deficiency in an enzyme that degrades this metabolite. The situation is frequently more complex: a biosynthetic enzyme may be abnormally regulated, a biosynthetic enzyme utilizing precursors of the metabolite may be deficient, a degradative enzyme several steps removed from the accumulated metabolite in the metabolic pathway may be deficient, or transport systems for the metabolite may be abnormal. Thus, investigations in search of altered enzymatic catalysis and/or transport should follow the identification of an abnormally accumulated metabolite wherever possible. Demonstration of a causal relationship between the accumulated metabolite and the neurological disorder is frequently difficult. Is the accumulation of the metabolite invariably associated with a stereotyped neurological dysfunction?

PYROGLUTAMIC ACID AND GLUTATHIONE Pyroglutamic aciduria was first described by Eldjarn and co-workers (81) in a teenage Norwegian boy who exhibited mental retardation, spastic tetraparesis, cerebellar intention tremor, ataxic gait, and dysarthria. Studies on cultured fibroblasts from this patient showed normal levels of the enzyme 5-oxoprolinase, which catalyzes the conversion of pyroglutamic acid to glutamate in an ATP-dependent reaction (82). Two young Swedish sisters, both of whom exhibited pyroglutamic aciduria (83), showed a deficiency of glutathione synthetase (84). More recently, two brothers with pyroglutamic aciduria were treated with vitamin E, which improved the phagocytic functions of their polymorphonuclear leukocytes (85). These brothers also had a deficiency of glutathione synthetase.

Meister and colleagues (84, 86) have suggested that a deficiency of glutathione synthetase short-circuits the γ-glutamyl cycle and leads to an overproduction of pyroglutamic acid. They propose that diminished glutathione synthesis permits an overproduction of γ-glutamyl cysteine, since glutathione is a feedback inhibitor of γ-glutamyl cysteine synthetase. Studies on cultured fibroblasts from a patient with glutathione deficiency support this proposal (87). Inhibition of γ-glutamyl transpeptidase by serine and borate increased cellular glutathione levels and decreased pyroglutamic acid synthesis. Variable results have been obtained from measurements of amino acid levels in erythrocytes from patients with pyroglutamic aciduria (88, 89). Majerus and colleagues (90) described an adult man with a red blood cell deficiency of glutathione synthetase associated with a hemolytic anemia. This patient showed no signs of neurological dysfunction; thus multiple forms of glutathione synthetase deficiency exist (91, 92).

Three patients with glutathionuria and mental retardation have been reported (93–96). Two of the patients have a deficiency of serum γ-glutamyl transpeptidase. The clearance of urinary amino acids was normal in two patients, which indicates that the reabsorption of amino acids in the renal proximal tubule was normal (95, 97). In addition, studies on red blood cells and cultured fibroblasts from one patient showed a deficiency of γ-glutamyl transpeptidase, but no abnormalities in the transport of amino acids (98–100). These data support a degradative role of γ-glutamyl transpeptidase and show that a deficiency of the enzyme is not associated with a disorder of amino acid transport—evidence against the theory that γ-glutamyl transpeptidase functions in the group translocation of amino acids.

Glutathione may exist in two forms where the sulfhydryl group of cysteine is either reduced (GSH) or oxidized (GSSG). A deficiency of glutathione reductase is associated with mental retardation and spasticity (101–104). Significantly increased levels of glutathione reductase in the serum of patients with cystic fibrosis has been reported (105). A deficiency of glutathione peroxidase in leukocytes is associated with a disorder of phagocytosis, but no neurological dysfunction has been recognized (106). Diminished levels of glutathione peroxidase have been found in the leukocytes and livers of rats fed a selenium-deficient diet (107, 108). Selenium deficiency in a variety of animals has been associated with several forms of muscular dystrophy syndrome (109, 110). In contrast, increased levels of glutathione peroxidase in red blood cells have been reported in patients with Down's syndrome (111).

A deficiency of the first enzyme in glutathione synthesis, γ-glutamyl cysteine synthetase, is associated with a spinocerebellar syndrome consisting of ataxia, dysmetria, diminished vibration and position senses, and hypoactive deep tendon reflexes (112–114). Transfer of the γ-glutamyl group of glutathione to an acceptor peptide or amino acid generates a dipeptide, cysteinyl glycine. Cysteinyl glycinuria has been described in a mentally retarded child who presumably lacks the necessary dipeptidase for catalyzing the degradation of cysteinyl glycine (115). Also of note is the occurrence of γ-glutamyl phenylalanine in the urine of newborn phenylketonurics (116). The significance of this observation is unclear at present.

PROLINE, HISTIDINE, AND GABA The foregoing array of disorders suggests that the conversion of γ-glutamyl moieties to glutamate proceeds primarily through the cyclized derivative of glutamate, pyroglutamic acid. Two other cyclized derivatives of glutamate, L-proline and Δ'-pyrroline-5-carboxylic acid are shown in Figure 2 (117, 118). Increased levels of

L-proline were found in the blood and urine of patients with deficiency of the enzyme proline oxidase, which has been designated type I hyperprolinemia. In type II hyperprolinemia, patients excrete an excess of both L-proline and Δ^1-pyrroline-5-carboxylic acid (119). This disorder appears to be associated with a deficiency of Δ^1-pyrroline-5-carboxylic acid dehydrogenase at the second step of proline catabolism.

Similarly, histidine catabolism also forms glutamate; a block in the primary pathway of histidine degradation results in the accumulation of histidine in blood, urine, and CSF (120). In some patients the levels of glutamate and glutamine were diminished in CSF. Histidinemia apparently results from a deficiency of histidase, which is the first enzyme in histidine catabolism and which catalyzes the deamination of L-histidine to form urocanic acid. Moderate mental retardation occurs in about two thirds of these patients and many have a defect in speech development that has been attributed to a short auditory memory span. No abnormality in auditory function has been reported.

The condensation of histidine with GABA results in the formation of the dipeptide γ-aminobutyryl-L-histidine, termed homocarnosine. Both homocarnosine and carnosine have been found in high concentrations in the olfactory bulbs of mammals and may function as possible neurotransmitters in olfaction (121). Elevated levels of homocarnosine were reported in the CSF of a mentally retarded woman with a progressive spastic tetraparesis and incontinence (122).

GABA is formed from glutamate in a decarboxylation reaction catalyzed by the enzyme glutamate decarboxylase and the cofactor pyridoxal phosphate. Measurements of glutamate decarboxylase activity in a kidney biopsy from a patient with pyridoxine (vitamin B_6)-dependent seizures suggest that the affinity of the enzyme for pyridoxal phosphate may be abnormal (123). Such an abnormality is consistent with the observation that infants with this seizure disorder respond to large doses of vitamin B_6. Administration of GABA to stop convulsions has also been attempted, but convincing studies of its efficacy are lacking (124).

Reports from several laboratories suggest that a deficiency of glutamate decarboxylase may be associated with Huntington's chorea (125, 126). Autopsy specimens from patients with this disorder show low levels of GABA and increased levels of γ-hydroxybutyric acid in the basal ganglia (125–128). It is of interest that γ-hydroxybutyric acid is a potent anesthetic agent (129). Membranes from the cerebellar cortex of patients with Huntington's chorea have an increased affinity for radiolabelled GABA (130). Further studies are required to determine whether or not GABA deficiency is a causal factor in the disease or merely a secondary phenomenon (131). Recently kainic acid, a naturally occurring glutamate analogue, was injected into the neostriatum of rats to produce degenerations resembling those found in Huntington's chorea (132). Large quantities of GABA were found in the urine of one male infant with hyper-β-alaninemia, who exhibited somnolence and generalized seizures (133). The presumptive cause for the accumulation of GABA and β-alanine in this patient is a deficiency of a common aminotransferase that catalyzes the degradation of GABA to succinic semialdehyde and β-alanine to malonic semialdehyde.

GLUTAMINE, AMMONIA, AND GLUTAMATE In contrast to many of the disorders described above that are apparently not exacerbated by the intake of nitrogenous foodstuffs, patients with several other disorders of glutamate metabolism cannot tolerate high nitrogen diets. A group of patients in Finland suffer from hereditary protein intolerance, which is associated with diminished intellectual function, progressive cortical atrophy, and spasticity (134). Measurements on leukocytes from one such patient suggest a deficiency of glutaminase may be responsible for this disorder (135). Another disorder in which nitrogenous foodstuffs such as protein cause illness is carbamyl phosphate synthetase (CPS) deficiency (136, 137). In mammalian liver, two forms of the enzyme exist. CPS I, a mitochondrial enzyme, participates in the fixation of ammonia via the urea cycle and is specifically activated by N-acetylglutamate. CPS II, a cytoplasmic enzyme, catalyzes the formation of carbamyl phosphate from CO_2, ATP, and the amide group of glutamate in pyrimidine biosynthesis. In several cases of CPS deficiency, measurements of the individual isoenzymes suggest that CPS I is consistently diminished, while CPS II may be decreased or even increased (137, 138). CPS deficiency is associated with seizures, mental retardation, and coma (136).

Both hereditary protein intolerance and CPS deficiency are associated with hyperammonemia. The toxic effects of ammonia on the nervous system have been recognized for many years from studies on hepatic encephalopathy (138–140). These patients exhibit asterixis, lethargy, confusion, and coma. Repeated bouts of hepatic encephalopathy appear to be associated with hepatocerebral degeneration consisting of dementia, dysarthria, ataxia, cerebellar intention tremor, choreiform movements, and dystonic posturing (141). The CSF of patients with hepatic encephalopathy contains significantly increased amounts of α-ketoglutaramate and L-glutamine (142, 143). Many patients with compromised liver function develop clinical signs of hepatic encephalopathy when presented with a load of ammonium ions (144). Recent studies suggest that branched chain amino acids given orally may provide effective therapy for chronic hepatic encephalopathy (145).

Both ammonium ions and glutamate in large doses are CNS toxins (137, 141, 146, 147). As noted above, exogenously administered ammonium ions can cause CNS dysfunction similar to a variety of syndromes associated with hyperammonemia (Table 3). The hyperammonemia associated with most of these disorders can be controlled by restricting the intake of nitrogen. High levels of blood ammonia in most of these syndromes are associated with vomiting, lethargy, seizures, and coma. The first five disorders in Table 3 are associated with blockage of the urea cycle, which is a major route for ammonia detoxification and elimination (137, 138). Hereditary protein intolerance and compromised hepatic function have been discussed above. In the remaining hyperammonemic syndromes, the mechanisms responsible for hyperammonemia are less clearly understood (146).

It is noteworthy that the same neuropathological changes in brain, i.e. Alzheimer's type II cells, are seen in almost all hyperammonemic syndromes (141). No histopathological differences distinguish acquired from congenital hyperammonemic states (146, 148).

In contrast to many of the hyperammonemic syndromes associated with inborn errors of metabolism and frequently manifesting neurological dysfunction in infancy or early childhood, intolerance to large amounts of glutamate has been observed in both children and adults. The molecular basis of intolerance for glutamate is not understood; it is not known if it is congenital or acquired. In adults this intolerance has been termed "Chinese restaurant syndrome" because it is frequently observed following consumption of a Chinese meal that contained large amounts of monosodium glutamate. The rather stereotyped presentation includes a burning sensation over the chest, neck, shoulders, and arms, as well as feelings of tightness and pressure over the chest and flushing of the malar areas of the face (149). Some psychiatric symptoms have been attributed to monosodium glutamate intake but further studies are needed to clarify these observations (150).

A few children who consume foods containing monosodium glutamate have been reported to develop "shivering" or "shuddering" attacks (151, 152). These attacks may be an expression of essential tremor due to a defect

Table 3 Hyperammonemic syndromes

1. Carbamyl phosphate synthetase deficiency
2. Ornithine carbamyl transferase deficiency
3. Arginino succinate synthetase deficiency (citrullinemia)
4. Arginino succinase deficiency (arginino succinic aciduria)
5. Arginase deficiency (hyperargininemia)
6. Hereditary protein intolerance
7. Hepatic insufficiency and encephalopathy
8. Reye's syndrome
9. Hyperlysinemia with lysine-induced crisis
10. Congenital lysine intolerance
11. Lysinuric protein intolerance
12. Hyperornithinemia associated with homocitrullinuria
13. Vitamine B_{12} unresponsive methylmalonic acidemia

in the immature brain. Also of note is the report of a mentally retarded male child who had a three- to five-fold elevation of glutamic acid in the CSF (153). The CSF glutamine and the blood ammonia were both normal in this child. The disorder may be sex-linked, since four other males in the family were mentally retarded.

Glutamate administered to experimental animals in large doses, orally or by direct application to the brain, causes structural damage in the CNS (7, 147, 154, 155), but the molecular basis of these nervous system lesions remains unknown. In contrast, diminished levels of glutamate have been found in the cerebella of both weaver and staggerer mice whose cerebella are deficient in granule cells (156, 157).

Summary

Disorders of glutamate metabolism are associated with profound alterations of CNS function. Although the precise biochemical mechanisms responsible for nervous system dysfunction in most of these disorders are unknown, it can be concluded that biochemical aberrations in the metabolism of glutamate usually result in malfunction of the CNS. As with most disorders of amino acid metabolism that are associated with brain dysfunction, the molecular mechanisms underlying the pathogenesis of these disorders remain obscure. Why do some biochemically affected individuals remain free of significant neurological involvement? The explanation that the accumulation of a specific metabolite is responsible for the pathogenesis of a disease process does not answer this and may frequently prove to be incorrect. More complete descriptions of these disorders of glutamate metabolism and their attendant neurological dysfunction will be required before any firm linkage between the biochemical defects and the clinical manifestations of the disease processes can be established. In addition, a much greater knowledge of the control mechanisms regulating glutamate metabolism in the CNS will be needed before the complexities of these disorders can be fully understood. Finally, the dual role of glutamate metabolites as intermediates in crucial pathways of metabolism and as neurotransmitters modulating electrical signals stresses the complexity and importance of glutamate metabolism in CNS function.

ACKNOWLEDGMENT

The author wishes to extend a special thank you to Drs. Curtis Morris and Louis Sokoloff for many helpful criticisms and suggestions. The author also thanks Deborah Downey for editorial assistance and Lorrie Gallagher for expert typing. This research was supported in part by a grant from the NIH NS 14543-02. The author is an investigator of the Howard Hughes Medical Institute.

Literature Cited

1. Waelsch, H. 1951. Glutamic acid and cerebral function. *Adv. Protein Chem.* 6:299–341
2. Gaull, G. E., Tallan, H. H., Lajtha, A., Rassin, D. K. 1975. Pathogenesis of brain dysfunction in inborn errors of amino acid metabolism. In *Biology of Brain Dysfunction,* ed. G. E. Gaull, 3:47–143. New York & London: Plenum
3. Prusiner, S. 1974. Regulation of allosteric enzymes in glutamine metabolism. In *The Neurosciences Third Study Program,* ed. F. O. Schmidt, F. G. Warden, pp. 853–61. Cambridge: MIT Press
4. Scriver, C. R., Rosenberg, L. E., eds. 1973. *Amino Acid Metabolism and Its Disorders.* Philadelphia: Saunders. 491 pp.
5. Meister, A. 1974. The γ-glutamyl cycle. Diseases associated with specific enzyme deficiencies. *Ann. Intern. Med.* 81:247–53
6. Meister, A., Tate, S. S. 1976. Glutathione and related γ-glutamyl compounds: biosynthesis and utilization. *Ann. Rev. Biochem.* 45:559–604
7. Filer, L. J. Jr., Garattini, S., Kare, M. R., Reynolds, W. A., Wurtman, R. J., eds. 1979. *Glutamic Acid: Advances in Biochemistry and Physiology,* pp. 1–395. New York: Raven
8. McGeer, E. G., Olney, J. W., McGeer, P. L., eds. 1978. *Kainic Acid as a Tool in Neurobiology,* pp. 1–267. New York: Raven
9. Krogsgaard-Larsen, P., Scheel-Krüger, J., Kofod, H., eds. 1979. *GABA-Neurotransmitters: Pharmacochemical, Biochemical and Pharmacological Aspects,* pp. 1–549. New York: Academic
10. Prusiner, S., Stadtman, E. R., eds. 1973. *The Enzymes of Glutamine Metabolism.* New York: Academic. 615 pp.
11. Reed, L. J. 1974. Multienzyme complexes. *Acc. Chem. Res.* 7:40–46
12. Frieden, C. 1976. The regulation of glutamate dehydrogenase. In *The Urea Cycle,* ed. S. Grisolia, R. Baguena, F. Mayor, pp. 59–71. New York: Wiley-Intersc. Publ.
13. Eisenberg, H. 1971. Glutamate dehydrogenase: anatomy of a regulatory enzyme. *Acc. Chem. Res.* 4:379–85
14. Holden, J. T., ed. 1962. *Amino Acid Pools,* pp. 472–78. Amsterdam: Elsevier
15. Prusiner, S., Stadtman, E. R. 1976. Regulation of glutaminase B in *Escherichia coli.* III. Control by nucleotides and divalent cations. *J. Biol. Chem.* 251:3463–69
16. Adler, S. P., Purich, D., Stadtman, E. R. 1975. Cascade control of *Escherichia coli* glutamine synthetase: properties of the P_{II} regulatory protein and the uridylyltransferase-uridylyl-removing enzyme. *J. Biol. Chem.* 250:6264–72
17. Stadtman, E. R., Ginsburg, A. 1974. The glutamine synthetase of *Escherichia coli:* structure and control. *The Enzymes* X:755–807
18. Deuel, T. G., Lerner, A., Albrycht, D. 1973. Regulation of glutamine synthetase from rat liver and rat kidney. See Ref. 10 pp. 129–44
19. Tate, S. S., Leu, F.-Y., Meister, A. 1972. Rat liver glutamine synthetase: preparation, properties, and mechanism of inhibition by carbamyl phosphate. *J. Biol. Chem.* 247:5312–21
20. Kingdon, H. S., Stadtman, E. R. 1967. Regulation of glutamine synthetase. X. Effect of growth conditions on the susceptibility of *Escherichia coli* glutamine synthetase to feedback inhibition. *J. Bacteriol.* 94:949–57
21. Prusiner, S., Miller, R. E., Valentine, R. C. 1972. Adenosine 3': 5'-cyclic monophosphate control of the enzymes of glutamine metabolism in *Escherichia coli. Proc. Natl. Acad. Sci. USA* 69:2922–26
22. Gaillardin, C. M., Magasanik, B. 1978. Involvement of the product of the *gln F* gene in autogenous regulation of glutamine synthetase formation in *Klebsiella aerogenes. J. Bacteriol.* 133:1329–38
23. Tyler, B. 1978. Regulation of the assimilation of nitrogen compounds. *Ann. Rev. Biochem.* 47:1127–62
24. Magasanik, B., Prival, M. J., Brenchley, J. E., Tyler, B. M., DeLeo, A. B., Streicher, S. L., Bender, R. A., Paris, C. G. 1974. Glutamine synthetase as a regulator of enzyme synthesis. *Curr. Top. Cell. Regulat.* 8:119–38
25. Streicher, S. L., DeLeo, A. B., Magasanik, B. 1976. Regulation of enzyme formation in *Klebsiella aerogenes* by episomal glutamine synthetase of *Escherichia coli. J. Bacteriol.* 127:184–92
26. Foor, F., Janssen, K., Magasanik, B. 1975. Regulation of synthesis of glutamine synthetase by adenylated glutamine synthetase. *Proc. Natl. Acad. Sci. USA* 72:4844–48
27. Reysset, B., Bott, K. F., Aubert, J.-P. 1978. Relationship between glutamine synthetase and sporulation. *Spores* VII:271–76

28. Resnick, A. D., Magasanik, B. 1976. L-asparaginase of *Klebsiella aerogenes. J. Biol. Chem.* 251:2722–28
29. Covarrubias, A. A., Sanchez-Pescador, R., Osorio, F., Bastarrachea, F. 1980. ColE1 hybrid plasmids containing *Escherichia coli* genes involved in the biosynthesis of glutamate and glutamine. *Plasmid* 3:150–64
30. Prusiner, S., Stadtman, E. R. 1976. Regulation of glutaminase B in *Escherichia coli.* II. Modulation of activity of carboxylate and borate ions. *J. Biol. Chem.* 251:3457–62
31. Berl, S., Clarke, D. D. 1970. Compartmentation of amino acid metabolism. *Handb. Neurochem.* 3:447–73
32. Bell, K. P., Norenburg, M. D. 1977. Glutamine synthetase: glial localization in brain. *Science* 195:1356–58
33. Bloom, F. E. 1972. Amino acids and polypeptides in neuronal function. *Neurosci. Res. Prog. Bull.* 10:122–251
34. Davidson, N. 1976. *Neurotransmitter Amino Acids.* London: Academic. 179 pp.
35. Johnson, J. L. 1972. Glutamic acid as a synaptic transmitter in the nervous system. A review. *Brain Res.* 37:1–19
36. Talman, W. T., Perrone, M. H., Reis, D. J. 1980. Evidence for L-glutamate as the neurotransmitter of baroreceptor afferent nerve fibers. *Science* 209: 813–15
37. Shank, R. P., Graham, L. T. Jr. 1978. The multiple roles of glutamate and aspartate in neural tissues. *Adv. Neurochem.* 3:165–201
38. Sharif, N. A., Roberts, P. J. 1980. Problems associated with the binding of L-glutamic acid to synaptic membranes: methodological aspects. *J. Neurochem.* 34:779–84
39. Roberts, E., Chase, T., Towers, D., eds. 1976. *GABA in Nervous System Function.* New York: Raven. 576 pp.
40. Alger, B. E., Nicoll, R. A. 1979. GABA-mediated biphasic inhibitory responses in hippocampus. *Nature* 281:315–17
41. Squires, R. F., Braestrup, C. 1977. Benzodiazepine receptors in rat brain. *Nature* 266:732–34
42. Möhler, H., Okada, T. 1977. Benzodiazepine receptor: demonstration in the central nervous system. *Science* 198:849–51
43. Tallman, J. F., Paul, S. M., Skolnick, P., Gallager, D. W. 1980. Receptors for the age of anxiety: pharmacology of the benzodiazepines. *Science* 207:274–81
44. MacDonald, J. F., Barker, J. L., Paul, S. M., Marangos, P. J., Skolnick, P. 1979. Inosine may be an endogenous ligand for benzodiazepine receptors on culture neurons. *Science* 205:715–17
45. Skolnick, P., Syapin, P. J., Paugh, B. A., Paul, S. M. 1979. Reduction in benzodiazepine receptors associated with Purkinje cell degeneration in "nervous" mutant mice. *Nature* 277:397–99
46. Eldjarn, L., Jellum, E., Stokke, O. 1974. Pyroglutamic acidemia. In *Heritable Disorders of Amino Acid Metabolism: Patterns of Clinical Expression and Genetic Variation,* ed. W. L. Nyhan, pp. 479–500. New York: Wiley
47. Van Der Werf, P., Orlowski, M., Meister, A. 1971. Enzymatic conversion of 5-oxo-L-proline (L-pyrrolidone carboxylate) to L-glutamate coupled with cleavage of adenosine triphosphate to adenosine dephosphate, a reaction in the γ-glutamyl cycle. *Proc. Natl. Acad. Sci. USA* 68:2982–85
48. Orlowski, M., Meister, A. 1973. γ-Glutamyl cyclotransferase: distribution, isoenzymic forms and specificity. *J. Biol. Chem.* 248:2836–44
49. Pisano, J. J. 1969. Peptides. *Handb. Neurochem.* I:53–74
50. Reichelt, K. L. 1970. The isolation of gamma-glutamyl peptides from monkey brain. *J. Neurochem.* 17:19–25
51. Tate, S. S., Meister, A. 1974. Interaction of γ-glutamyl transpeptidase with amino acids, dipeptides, and derivatives and analogs of glutathione. *J. Biol. Chem.* 249:7593–7602
52. Silbernagel, S., Pfaller, W., Heinle, H., Wendel, A. 1978. Topology and function of renal γ-glutamyl transpeptidase. In *Functions of Glutathione in Liver and Kidney,* ed. H. Sies, A. Wendel, pp. 60–69. Berlin: Springer
53. Curthoys, N. P., Hughey, R. P. 1979. Characterization and physiological function of rat renal γ-glutamyl transpeptidase. *Enzyme* 24:383–403
54. McIntyre, T. M., Curthoys, N. P. 1979. Comparison of the hydrolytic and transfer activities of rat renal γ-glutamyl transpeptidase. *J. Biol. Chem.* 254:6499–6504
55. Hahn, R., Wendel, A., Flohé, L. 1978. The fate of extracellular glutathione in the rat. *Biochim. Biophys. Acta* 539: 324–37
56. Inoue, M., Horiuchi, S., Morino, Y. 1977. γ-Glutamyl transpeptidase in rat ascites tumor cell LY-5: lack of functional correlation of its catalytic activity

with the amino acid transport. *Eur. J. Biochem.* 78:609–15

57. Wendel, A., Heinle, H., Silbernagel, S. 1978. The degradation of glutathione derivatives in the rat kidney. In *Biochemical Nephrology*, ed. W. G. Guder, U. Schmidt, pp. 73–84. Bern: Huber
58. Ball, E. G., Cooper, O., Clarke, E. C. 1953. On the hydrolysis and transpeptidation of glutathione in marine forms. *Biol. Bull.* 105:369–70
59. Binkley, F. 1954. Discussion, glutathione. *Proc. Symp. at Ridgefield, Conn.*, ed. S. Colowick et al, pp. 160–61. New York: Academic
60. Orlowski, M., Meister, A. 1970. The γ-glutamyl cycle: a possible transport system for amino acids. *Proc. Natl. Acad. Sci. USA* 67:1248–55
61. Griffith, O. W., Bridges, R. J., Meister, A. 1979. Transport of γ-glutamyl amino acids: role of glutathione and γ-glutamyl transpeptidase. *Proc. Natl. Acad. Sci. USA* 76:6319–22
62. Sikka, S. C., Kalra, V. K. 1980. γ-Glutamyl transpeptidase-mediated transport of amino acid in lecithin vesicles. *J. Biol. Chem.* 255:4399–4402
63. Prusiner, S., Doak, C. W., Kirk, G. 1976. A novel mechanism for group translocation: substrate-product reutilization by γ-glutamyl transpeptidase in peptide and amino acid transport. *J. Cell Physiol.* 89:853–64
64. Flohé, L., Benöhr, H. Ch., Sies, H., Waller, H. D., Wendel, A., eds. 1974. Glutathione. *Proc. 16th Conf. German Soc. Biol. Chem., Tubingen, March 1973.* Stuttgart: Thieme. 316 pp.
65. Griffith, O. W., Meister, A. 1979. Translocation of intracellular glutathione to membrane-bound γ-glutamyl transpeptidase as a discrete step in the γ-glutamyl cycle: glutathionuria after inhibition of transpeptidase. *Proc. Natl. Acad. Sci. USA* 76:268–72
66. Prusiner, P. E., Prusiner, S. B. 1978. Modulation of γ-glutamyl transpeptidase activity from bovine choroid plexus. *J. Neurochem.* 30:1261–67
67. Mamelok, R. D., Groth, D. F., Prusiner, S. B. 1980. Separation of membrane-bound γ-glutamyl transpeptidase from brush border transport and enzyme activities. *Biochemistry* 19:2367–73
68. Pocius, P. A., Baumrucker, C. R., McNamara, J. P., Bauman, D. E. 1980. γ-Glutamyl transpeptidase in rat mammary tissue: activity during lactogenesis and regulation by prolactin. *Biochem. J.* 188:565–68
69. DeBault, L. E., Cancilla, P. A. 1980. γ-Glutamyl transpeptidase in isolated brain endothelial cells: induction by glial cells *in vitro*. *Science* 207:653–55
70. Betz, A. L., Firth, J. A., Goldstein, G. W. 1980. Polarity of the blood-brain barrier: distribution of enzymes between the luminal and antiluminal membranes of brain capillary endothelial cells. *Brain Res.* 192:17–28
71. Jones, D. P., Moldeus, P., Stead, A. E., Ormstad, K., Jornvall, H., Orrenius, S. 1979. Metabolism of glutathione and a glutathione conjugate by isolated kidney cells. *J. Biol. Chem.* 254:2787–92
72. Ormstad, K., Orrenius, S. 1980. The localization of renal glutathione oxidase activity studied in the isolated, perfused rat kidney. *Biochem. Biophys. Res. Commun.* 92:540–45
73. Griffith, O. W., Tate, S. S. 1980. The apparent glutathione oxidase activity of γ-glutamyl transpeptidase: chemical mechanism. *J. Biol. Chem.* 255:5011–14
74. Arias, I. M., Jakoby, W. B., eds. 1976. *Glutathione: Metabolism and Function*, pp. 1–381. New York: Raven
75. Apontweil, P., Berends, W. 1975. Isolation and initial characterization of glutathione-deficient mutants of *Escherichia coli* K 12. *Biochim. Biophys. Acta* 399:10–22
76. Cooper, A. J. L., Meister, A. 1973. Glutamine transaminases from liver and kidney. See Ref. 10, pp. 207–226
77. Hersh, L. B. 1972. Rat liver ω-amidase kinetic evidence for an acylenzyme intermediate. *Biochemistry* 11:2251–56
78. Folk, J. E., Finlayson, J. S. 1977. The epsilon-(gamma-glutamyl)lysine crosslink and the catalytic role of the transglutaminases. *Adv. Protein Chem.* 31:1–133
79. Davies, P. J. A., Davies, D. R., Levitzki, A., Maxfield, F. R., Milhaud, P., Willingham, M. C., Pastan, I. H. 1980. Transglutaminase is essential in receptor mediated endocytosis of α_2 macroglobulin and polypeptide hormones. *Nature* 283:162–67
80. Folk, J. E., Park, M. H., Chung, S. I., Schrode, J., Lester, E. P., Cooper, H. L. 1980. Polyamines as physiological substrates for transglutaminases. *J. Biol. Chem.* 255:3695–3700

81. Jellum, E., Kluge, T., Börresen, H. C., Stokke, O., Eldjarn, L. 1970. Pyroglutamic aciduria—a new inborn error of metabolism. *Scand. J. Clin. Lab. Invest.* 26:327–35
82. Strömme, J. H., Eldjarn, L. 1972. The metabolism of L-pyroglutamic acid in fibroblasts from a patient with pyroglutamic aciduria: the demonstration of an L-pyroglutamate hydrolase system. *Scand. J. Clin. Lab. Invest.* 29: 335–42
83. Hagenfeldt, L., Larsson, A., Zetterström, R. 1974. Pyroglutamic aciduria. *Acta Paediatr. Scand.* 63:1–8
84. Wellner, V. P., Sekura, R., Meister, A., Larsson, A. 1974. Glutathione synthetase deficiency, an inborn error of metabolism involving the γ-glutamyl cycle in patients with 5-oxoprolinuria (pyroglutamic aciduria). *Proc. Natl. Acad. Sci. USA* 71:2505–9
85. Boxer, L., Oliver, J. M., Spielberg, S. P., Allen, J. M., Schulman, J. D. 1979. Protection of granulocytes by vitamin E in glutathione synthetase deficiency. *N. Engl. J. Med.* 301:901–5
86. Larsson, A., Mattsson, B. 1976. On the mechanism of 5-oxoproline overproduction in 5-oxoprolinuria. *Clin. Chim. Acta* 67:245–53
87. Spielberg, S. P., Butler, J. DeB., MacDermott, K., Schulman, J. D. 1979. Treatment of glutathione synthetase deficient fibroblasts by inhibiting γ-glutamyl transpeptidase activity with serine and borate. *Biochem. Biophys. Res. Commun.* 89:504–11
88. Marstein, S., Jellum, E., Halpern, B., Eldjarn, L., Perry, T. L. 1976. Biochemical studies of erythrocytes in a patient with pyroglutamic acidemia (5-oxoprolinemia). *N. Engl. J. Med.* 295:406–12
89. Hagenfeldt, L., Larsson, A., Andersson, R. 1978. The γ-glutamyl cycle and amino acid transport: studies of free amino acids, γ-glutamylcysteine and glutathione in erythrocytes from patients with 5-oxoprolinuria (glutathione synthetase deficiency). *N. Engl. J. Med.* 299:587–90
90. Mohler, D. N., Majerus, P. W., Minnich, V., Hess, C. E., Garrick, M. D. 1970. Glutathione synthetase deficiency as a cause of hereditary hemolytic disease. *N. Engl. J. Med.* 283: 1253–56
91. Spielberg, S. P., Kramer, L. I., Goodman, S. I., Butler, J., Tietze, F., Quinn, P., Schulman, J. D. 1977. 5-Oxoprolinuria: biochemical observations and case report. *J. Pediatr.* 91: 237–41
92. Spielberg, S. P., Garrick, M. D., Corash, L. M., Butler, J. DeB., Tietze, F., Rogers, L., Schulman, J. D. 1978. Biochemical heterogeneity in glutathione synthetase deficiency *J. Clin. Invest.* 61:1417–20
93. Goodman, S. I., Mace, J. W., Pollak, S. 1971. Serum gamma-glutamyl transpeptidase deficiency. *Lancet* I:234–35
94. O'Daly, S. 1968. An abnormal sulphydryl compound in urine. *Ir. J. Med. Sci. (Ser. VII)* 1:578–79
95. Wright, E. C., Stern, J., Ersser, R., Patrick, A. D. 1979. Glutathionuria: γ-glutamyl transpeptidase deficiency. *J. Inher. Metab. Dis.* 2:3–7
96. Griffith, O. W., Meister, A. 1980. Excretion of cysteine and γ-glutamylcysteine moieties in human and experimental animal γ-glutamyl transpeptidase deficiency. *Proc. Natl. Acad. Sci. USA* 77:3384–87
97. Schulman, J. D., Goodman, S. I., Mace, J. W., Patrick, A. D., Tietze, F., Butler, E. J. 1975. Glutathionuria: inborn error of metabolism due to tissue deficiency of gamma-glutamyl transpeptidase. *Biochem. Biophys. Res. Commun.* 65:68–74
98. Pellefigue, F., Butler, J. DeB., Spielberg, S. P., Hollenberg, M. D., Goodman, S. I., Schulman, J. D. 1976. Normal amino acid uptake by cultured human fibroblasts does not require gamma-glutamyl transpeptidase. *Biochem. Biophys. Res. Commun.* 73:997–1002
99. Beutler, E. 1976. Glutathione deficiency, pyroglutamic acidemia and amino acid transport. *N. Engl. J. Med.* 295:441–43
100. Marstein, S., Jellum, E., Halpern, B., Eldjarn, L., Perry, T. L. 1976. Biochemical studies of erythrocytes in a patient with pyroglutamic acidemia (5-oxoprolinemia). *N. Engl. J. Med.* 295:406–12
101. Waller, H. D., Löhr, G. W., Zysno, E., Gerok, W., Voss, D., Strauss, G. 1965. Glutathionreduktasemangel mit hämatologischen und neurologischen Störungen (autosomal dominant verebliche Bildung eines pathologischen Enzyms). *Klin. Wochenschr.* 43:413–27
102. Waller, H. D., Benöhr, H. Ch., Lehmann, W., Ericksson, A. W. 1972. Aktivierung der Glutathionreduktase roter Blutzellen mit Flavin-Adenin-Dinucleotid (FAD) (Untersuchungen an roter Blutzellen von Skoltlappen). *Klin. Wochenschr.* 50:462–66

103. Fajnholc, N. E., Kaminsky, E., Machtey, I., deVries, A. 1971. Hereditary erythrocyte glutathione reductase deficiency. *Rev. Eur. Etud. Clin. Biol.* 16:987–92

104. Beutler, E. 1975. Disorders in glutathione metabolism. *Life Sci.* 16:1499–1506

105. Shapiro, B. L., Smith, Q. T., Warwick, W. J. 1975. Serum glutathione reductase and cystic fibrosis. *Pediatr. Res.* 9:885–88

106. Necheles, T. F. 1974. The clinical spectrum of glutathione-peroxidase deficiency. In *Glutathione,* ed. L. Flohé, H. Ch. Benöhr, H. Sies, H. D. Waller, A. Wendel, pp. 173–80. Stuttgart: Thieme

107. Serfass, R. E., Ganther, H. E. 1975. Defective microbicidal activity in glutathione peroxidase-deficient neutrophils of selenium-deficient rats. *Nature* 255: 640–41

108. Lawrence, R. A., Burk, R. F. 1976. Glutathione peroxidase activity in selenium-deficient rat liver. *Biochem. Biophys. Res. Commun.* 71:952–58

109. Bains, B. S., MacKenzie, M. A., McKenzie, R. A. 1975. Selenium-responsive myopathy in broiler breeder hens in Queensland. *Aust. Vet. J.* 51:140–45

110. Stadtman, T. C. Selenium biochemistry. 1974. *Science* 183:915–22

111. Sinet, P. M., Michelson, A. M., Bazin, A., LeJeune, J., Jerome, H. 1975. Increase in glutathione peroxidase activity in erythrocytes from trisomy 21 subjects. *Biochem. Biophys. Res. Commun.* 67:910–15

112. Konrad, P. N., Richards, R., Valentine, W. N., Paglia, D. E. 1972. γ-Glutamylcysteine synthetase deficiency. A cause of hereditary hemolytic anemia. *N. Engl. J. Med.* 286:557–61

113. Richards, F., Cooper, M. R., Pearce, L. A., Cowan, R. J., Spurr, C. L. 1974. Familial spinocerebellar degeneration, hemolytic anemia and glutathione deficiency. *Arch. Intern. Med.* 134:534–37

114. Oshima, R. G., Rhead, W. J., Thoene, J. G., Schneider, J. A. 1976. Cystine metabolism in human fibroblasts. Comparison of normal cystinotic and γ-glutamylcysteine synthetase-deficient cells. *J. Biol. Chem.* 251:4287–93

115. Ben-Ami (Bodonyi), E., Burstein, I., Cohen, B. E., Szeinberg, A. 1973. Cysteine peptiduria in a mentally retarded patient. *Clin. Chim. Acta* 45:335–39

116. Peck, H., Pollitt, R. J. 1979. The occurrence of γ-glutamylphenyl-alanine in the urine of newborn phenylketonurics. *Clin. Chim. Acta* 94:237–40

117. Scriver, C. R., Efron, M. L. 1972. Disorders of proline and hydroxyproline metabolism. In *The Metabolic Basis of Inherited Disease,* ed. J. Stanbury, J. Wyngaarden, D. Fredrickson, pp. 351–69. New York: McGraw-Hill. 3rd ed.

118. Valle, D., Goodman, S. I., Applegarth, D. A., Shih, V. E., Phang, J. M. 1976. Type II hyperprolinemia. Δ-Pyrroline-5-carboxylic acid dehydrogenase deficiency in cultured skin fibroblasts and circulating lymphocytes. *J. Clin. Invest.* 58:598–603

119. Similä, S. 1979. Hydroxyproline metabolism in type II hyperprolineamia. *Ann. Clin. Biochem.* 16:177–81

120. La Du, B. 1972. Histidinemia. See Ref. 117, pp. 338–50

121. Margolis, F. L. 1974. Carnosine in the primary olfactory pathway. *Science* 184:909–911

122. Gjessing, L. R., Sjaastad, O. 1974. Homocarnosinosis: a new metabolic disorder associated with spasticity and mental retardation. *Lancet* II:1028

123. Yoshida, T., Tada, K., Arakawa, T. 1971. Vitamin B_6-dependency of glutamic acid decarboxylase in the kidney from a patient with vitamin B_6-dependent convulsion. *Tohoku J. Exp. Med.* 104:195–98

124. Tower, D. B. 1969. Neurochemical mechanisms. In *Basic Mechanisms of the Epilepsies,* ed. H. Jasper, A. Ward, A. Pope, pp. 611–38. Boston: Little, Brown

125. Perry, T. L., Hansen, S., Kloster, M. 1973. Huntington's chorea. Deficiency of γ-aminobutyric acid in brain. *N. Engl. J. Med.* 288:337–42

126. Enna, S. J., Bird, E. D., Bennet, J. P. Jr., Bylund, D. B., Yamamura, H. I., Iverson, L. L., Snyder, S. H. 1976. Huntington's chorea. Changes in neurotransmitter receptors in the brain. *N. Engl. J. Med.* 294:1305–9

127. McGeer, P. L., McGeer, E. G., Fibiger, H. C. 1973. Choline acetylase and glutamic acid decarboxylase in Huntington's chorea. *Neurology* 23:912–17

128. Ando, N., Gold, B. I., Roth, R. H. 1977. Regional distribution of γ-hydroxybutyrate in Huntington's chorea brains. *Trans. Am. Soc. Neurochem.* 8:270 (Abstr.)

129. Laborit, H. 1973. Gamma-hydroxybutyrate, succinic semialdehyde and sleep. *Prog. Neurobiol.* 1:255–74

130. Lloyd, K. G., Davidson, L. 1979. [^{3}H]GABA binding in brains from Huntington's chorea patients: altered regulation by phospholipids? *Science* 205:1147–49
131. Curzon, G., Kantamaneni, B. D. 1978. Transmitters and related substances in brain material: an index of terminal and/or post-mortem biochemical change. *Lancet* I:156–57
132. Coyle, J. T., McGeer, E. G., McGeer, P. L., Schwarcz, R. 1978. Neostriatal injections: a model for Huntington's chorea. See Ref. 8, pp. 139–60
133. Scriver, B., Pueschel, S., Davies, E. 1966. Hyper-β-alaninemia associated with β-aminobutyricaciduria, somnolence and seizures. *N. Engl. J. Med.* 274:636–43
134. Malmquist, J., Jagenburg, R., Lindstedt, G. 1971. Familial protein intolerance: possible nature of enzyme defect. *N. Engl. J. Med.* 284:997–1002
135. Malmquist, J., Hetter, B. 1970. Leucocyte glutaminase in familial protein intolerance. *Lancet* II:129–30
136. Freeman, J. M., Nicholson, J. F., Schimke, R. T., Rowland, L. P., Carter, S. 1970. Congenital hyperammonemia: association with hyperglycinemia and decreased levels of carbamyl phosphate synthetase. *Arch. Neurol.* 23:430–37
137. Grisolia, S., Baguena, R., Mayor, F., eds. 1976. *The Urea Cycle.* New York: Wiley-Intersci. Publ. 577 pp.
138. Shih, V. E. 1976. Hereditary urea-cycle disorders. See Ref. 137, pp. 367–414
139. Summerskill, W. H. J., Davidson, E. A., Sherlock, S., Steiner, R. E. 1956. The neuropsychiatric syndrome associated with hepatic cirrhosis and an extensive portal collateral circulation. *Q. J. Med. New Series XXV* 98:245–66
140. Walker, C. O., Schenker, S. 1970. Pathogenesis of hepatic encephalopathy with special reference to the role of ammonia. *Am. J. Clin. Nutr.* 23:619–32
141. Victor, M., Adams, R. D., Cole, M. 1965. The acquired (non-Wilsonian) type of chronic hepatocerebral degeneration. *Medicine* 44:345–93
142. Vergara, F., Plum, F., Duffy, T. E. 1974. α-Ketoglutaramate: increased concentrations in the cerebrospinal fluid of patients in hepatic coma. *Science* 183:81–83
143. Duffy, T. E., Vergara, F., Plum, F. 1974. α-Ketoglutaramate in hepatic encephalopathy. In *Brain Dysfunction in Metabolic Disorders,* ed. F. Plum, 53:39–52. New York: Raven Res. Publ. Assoc. Res. Nerv. Ment. Dis.
144. Stahl, J. 1963. Studies of the blood ammonia in liver disease. Its diagnostic, prognostic, and therapeutic significance. *Ann. Intern. Med.* 58:1–24
145. Freund, H., Yoshimura, N., Fischer, J. E. 1979. Chronic hepatic encephalopathy: long-term therapy with a branched-chain amino acid enriched elemental diet. *J. Am. Med. Assoc.* 242: 347–49
146. Hsia, Y. E. 1974. Inherited hyperammonemic syndromes. *Gastroenterology* 67:347–74
147. Olney, J. W., 1974. Toxic effects of glutamate and related amino acids on the developing central nervous system. See Ref. 46, pp. 501–12
148. Bruton, C. J., Corsellis, J. A. N., Russell, A. 1970. Hereditary hyperammonaemia. *Brain* 93:423–34
149. Schaumburg, H. H., Buck, R., Gerstle, R., Mashman, J. H. 1969. Monosodium L-glutamate: its pharmacology and role in the Chinese restaurant syndrome. *Science* 163:826–28
150. Colman, A. D. 1979. Possible psychiatric reactions to monosodium glutamate. *N. Engl. J. Med.* 300:503
151. Reif-Lehrer, L., Stemmermann, M. G. 1975. Monosodium glutamate intolerance in children. *N. Engl. J. Med.* 293:1204–5
152. Andermann, F., Vanasse, M., Wolfe, L. S. 1976. Shuddering attacks in children: essential tremor and monosodium glutamate. *N. Engl. J. Med.* 295: 174
153. Yoshida, T., Tada, K., Mizuno, T., Wada, Y., Akabane, J., Ogasawara, J., Minagawa, A., Morikawa, T., Okamura, T. 1964. A sex-linked disorder with mental and physical retardation characterized by cerebrocortical atrophy and increase of glutamic acid in the cerebrospinal fluid. *Tohoku J. Exp. Med.* 83:261–69
154. Krieger, D. T., Liotta, A. S., Nicholsen, G., Kizer, S. 1979. Brain ACTH and endorphin reduced in rats with monosodium glutamate-induced arcuate nuclear lesions. *Nature* 278: 562–63
155. Eskay, R. L., Brownstein, M. J., Long, R. T. 1979. α-Melanocyte-stimulating hormone: reduction in adult rat brain after monosodium glutamate treatment of neonates. *Science* 205:827–29
156. McBride, W. J., Aprison, M. H., Kusano, K. 1976. Contents of several amino acids in the cerebellum, brain stem and cerebrum of the "staggerer", "weaver" and "nervous" neuro-

logically mutant mice. *J. Neurochem.* 26:867–70

157. Roffler-Tarlov, S., Sidman, R. L. 1978. Concentrations of glutamic acid in the cerebellar cortex and deep nuclei of normal mice, weaver, staggerer, and nervous mice. *Brain Res.* 142:269–83
158. Perry, T. L., Hansen, S., Kennedy, J. 1975. CSF amino acids and plasma-CSF amino acid ratios in adults. *J. Neurochem.* 24:587–89
159. Perry, T. L., Hansen, S. 1976. Taurine content of epileptogenic foci in human brain. In *Taurine*, ed. R. Huxtable, A. Barbeau. New York: Raven
160. Perry, T. L., Hansen, S. 1976. Is GABA detectable in human CSF? *J. Neurochem.* 27:1537–38
161. Cooper, A. J. L., Dhar, A. K., Kutt, H., Duffy, T. E. 1977. GC assay of α-ketoglutaramic and 2-pyrrolidone-5-carboxylic acids in CSF. *Trans. Am. Soc. Neurochem.* 8:270 (Abstr.)

Ann. Rev. Med. 1981. 32:543–57
Copyright © 1981 by Annual Reviews Inc. All rights reserved

TESTICULAR CANCER

❖7418

Donald G. Skinner, M.D.

Division of Urology, University of Southern California School
of Medicine, 1200 North State Street, Los Angeles, California 90033

Peter T. Scardino, M.D.

Division of Urology, Baylor College of Medicine, Houston, Texas 77030

John R. Daniels, M.D.

Division of Medical Oncology, University of Southern California School of
Medicine, 1200 North State Street, Los Angeles, California 90033

INTRODUCTION

Dramatic advances have occurred in the therapy and survival of patients afflicted with all stages of germ cell tumors of the testis; in fact, these tumors now carry the best prognosis of any solid tumor occurring in man. Several developments are responsible for this improved survival.

First and most prominent has been the development and application of truly effective chemotherapy evolving from actinomycin D to the effective combination of vinblastine sulfate and bleomycin (16) to our currently most effective combination of platinum, vinblastine sulfate, and bleomycin (4).

Second, during this same era we demonstrated that a meticulous retroperitoneal lymph node dissection could remove all retroperitoneal nodes (9). Clinical reports from several institutions convincingly demonstrated that retroperitoneal recurrence following lymphadenectomy rarely, if ever, occurred and that postoperative radiotherapy was unnecessary and jeopardized subsequent effective use of the newer, more potent chemotherapeutic combination (3, 20, 22, 23).

Third was a clarification of the pathogenesis of these tumors leading to a classification that eliminated much of the confusing terminology (5, 6).

Fourth was the development of sensitive and specific radioimmunoassay

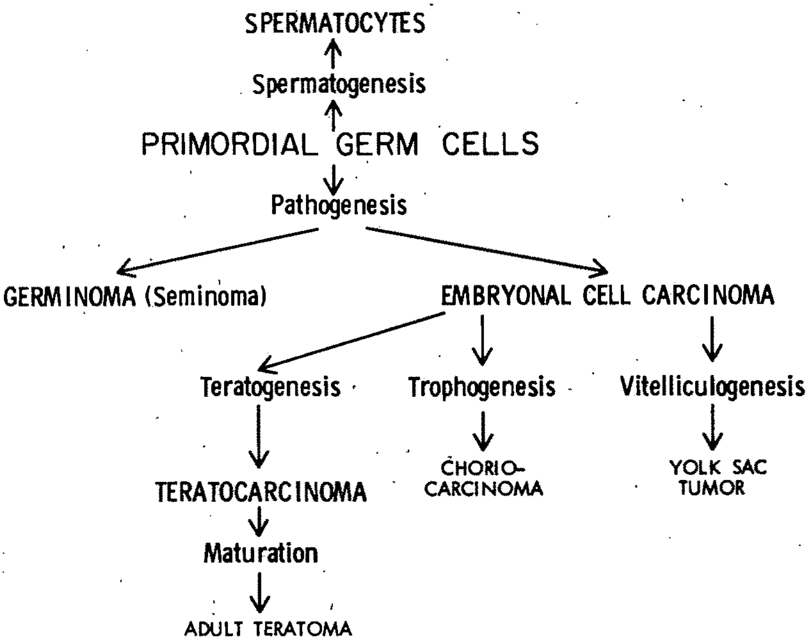
0066-4219/81/0401-0543$01.00

techniques to measure serum levels of human chorionic gonadotrophin (HCG) and alpha fetoprotein (AFP) (12, 19, 20). Production of HCG by syncytiotrophoblastic giant cells and of AFP by embryonal carcinoma cells has been demonstrated by specific immunohistochemical techniques (11). Therefore, the presence of an elevated serum level of these markers reliably indicates the presence of malignant elements, usually embryonal carcinoma, and the markers have assumed a prominent place in the management of germ cell tumors (11).

The integration and utilization of these four developments improved dramatically the poor prognosis formerly associated with the nonseminomatous varieties of germ cell tumors, so that by the end of the 1970s these tumors, along with the seminomatous variety of malignancy, carry the best prognosis of any genitourinary malignant tumor and of any solid tumor encountered in man.

CLASSIFICATION

Basic to the understanding of these tumors is a realization that all malignant germ cell tumors of the testis are derived from the primordial germ cells, and that pathogenesis may lead to the development of seminoma or embryonal carcinoma. Embryonal carcinoma is a unique pluripotential stem cell that can undergo vitelliculogenesis into yolk sac carcinoma, trophogenesis into choriocarcinoma, or, more commonly, teratogenesis into teratocarcinoma with ultimate further differentiation into teratoma (Figure 1) (5). Hence, it comes as no surprise that these tumors are often mixed and that a wide variety of cell types may be found in either the primary or metastatic foci. The natural history and biological behavior of tumors derived from embryonal carcinoma are similar regardless of the cell mix with the exception of the exceedingly rare pure choriocarcinoma. Therefore, classification can be simplified into pure seminoma vis à vis the heterogeneous pool of nonseminomatous or mixed tumors. It should be emphasized that the biologic behavior is determined by the nonseminomatous component; that tumors containing any amount of nonseminomatous tumor behave as a nonseminoma; that pure seminomas usually remain that way unless they harbor syncytiotrophoblastic giant cells or secrete HCG; that pure choriocarcinoma is an exceedingly rare tumor comprising less than 0.4% of all germinal tumors; and that choriocarcinoma mixed with other cell types behaves like any other mixed germinal tumor and does not carry the same ominous prognosis as a pure choriocarcinoma. Choriocarcinoma implies the combination of syncytiotrophoblasts surrounded by cytotrophoblasts arranged in a pseudovillus pattern. Unless all of these criteria are seen, the term should not be used.



(Friedman, 1976)

Figure 1 Pathways of differentiation and transformation of testis tumor.

EPIDEMIOLOGY

Germ cell tumors of the testis primarily affect men between the ages of 16 and 45 with a peak for seminoma in the mid 30s and for nonseminoma in the early 20s. The overall incidence is clearly on the increase for unknown reasons and testicular cancer comprises the most common cancer other than lymphoma encountered in men between the ages of 29 and 35 (24). There is an increased incidence in patients with cryptorchid testes not pexed before the age of six years. For unknown reasons, testis tumors rarely occur in the black population; caucasians have a four times higher incidence (8).

Most patients present with a painless mass noted in the testis, but nearly 20% will seek medical attention because of pain (24). Many of these patients are initially treated as having epididymitis. Ultrasound tests will often differentiate between epididymitis and a testicular tumor and should be obtained whenever any doubt of diagnosis exists (24). Any diagnostic suspicion of a testicular tumor requires exploration through an inguinal incision, with delivery of the testis from the scrotum with its tunics intact after the blood supply to the testis has been secured by a noncrushing vascular clamp. If a seminoma is encountered, careful histological evaluation of the entire testis is essential to rule out the possible coexistence of a small nonseminomatous element.

NATURAL HISTORY AND STAGING

Testis tumors predictably metastasize through lymphatics to the regional nodes located in the retroperitoneum in the periaortic and perivena caval regions just caudal to the renal hilum. Approximately 25% of patients with pure seminoma will have metastasized when first seen for therapy, compared to more than 66% of patients with nonseminomatous tumors who have metastasized at the time of initial presentation. Once the diagnosis of a germinal tumor of the testis has been established, diagnostic evaluation should include a chest x ray with full lung tomograms to rule out pulmonary metastatic disease and an IVP to detect bulk retroperitoneal disease. The hallmark of extensive retroperitoneal nodal disease is lateral deviation of the proximal third of the ureter. Bilateral pedal lymphangiography is usually obtained in patients with pure seminoma, but has not proven useful in patients with nonseminomatous tumors for four reasons. First, there is a high incidence of false negatives in patients with minimal disease; second, there is the discomfort and morbidity of the procedure; third, surgical staging is performed in each case regardless of the results of lymphangiogram unless massive abdominal disease is present; and, fourth, massive disease can be readily detected by careful abdominal palpation together with an IVP or CT scan of the abdomen, and that surgical staging renders the test unnecessary. The purpose of clinical staging in patients with nonseminomatous germinal tumors of the testis is to distinguish those with advanced disease who require initial chemotherapy from those with early disease who will be treated initially by surgery. If advanced disease is determined on the basis of full lung tomograms, IVP, CT scan, or physical exam, initial therapy, regardless of cell type, should be chemotherapy. If there is no evidence of advanced disease, patients with pure seminoma are managed by radiotherapy utilizing portals according to the results of lymphangiogram. Patients with nonseminoma are surgically staged by means of a retroperitoneal lymph node dissection. The staging scheme used in the United States is noted in Table 1.

MANAGEMENT OF SEMINOMA

Patients with clinical Stage A or B disease with normal postorchiectomy serum levels of HCG and AFP are managed by radiation therapy. If the lymphangiogram is within normal limits, 2500–3000 rads directed to the retroperitoneum and ipsilateral iliac area cures 92–98% of patients (2, 21,

Table 1 Staging scheme for testicular cancer

Stage A	Tumor confined to testis; no evidence of spread beyond confines of testis
Stage B_1	Evidence of minimal retroperitoneal lymph node metastases, detected by retroperitoneal lymph node dissection or lymphangiography (fewer than 6 positive nodes; no node greater than 2 cm in diameter)
Stage B_2	Evidence of moderate retroperitoneal lymph node spread (more than 6 positive nodes; or any node longer than 2 cm in diameter)
Stage B_3	Massive retroperitoneal lymph node involvement, but without evidence of spread above diaphragm or to solid visceral organs. Diagnosis made clinically on basis of palpable abdominal mass
Stage C	Metastatic tumor noted above diaphragm or involvement of solid visceral organs, brain, or bone

25). There appears to be little need in Stage A patients for extended portals to include the mediastinum or supraclavicular areas. If the lymphangiogram is positive (Stage B) but there is no evidence of bulk abdominal disease, the radiation portals are extended to include the mediastinum and supraclavicular areas as well as the retroperitoneum, with cure rates of 74–93% (21).

The management of advanced disease or patients with elevated serum HCG remains controversial. Standard radiation therapy in patients with advanced seminoma results in tumor-free survival of less than 40%, considerably less than is currently achieved in the more virulent nonseminomatous tumors of similar stage (2, 25). Seminoma is sensitive to chemotherapy with a variety of agents including alkylating agents and the combination of platinum, vinblastine sulfate, and bleomycin. Consequently, we treat these patients initially with 3–4 cycles of this combination chemotherapy followed by surgical resection of identifiable residual retroperitoneal or pulmonary disease. The histology of the resected tissue determines further chemotherapy. We reserve radiation therapy for highly selected individuals, usually those with unresectable mediastinal disease.

It is difficult to interpret HCG elevation in patients with seminoma. Mostofi (15) reported that 50% of patients whose primary seminoma contains syncytiotrophoblastic giant cells subsequently died of metastatic disease if treated with conventional radiation therapy. Immunocytochemical techniques have demonstrated HCG only within syncytiotrophoblastic giant cells, which indicates that elevated serum HCG levels reliably predict the presence of these cells within the tumor (11). Consequently, histologically pure seminoma, which is associated with elevated serum levels of HCG, should logically be treated as a nonseminomatous tumor using lymphadenectomy and chemotherapy according to pathologic stage.

MANAGEMENT OF NONSEMINOMA

Serum Tumor Markers

Serum biochemical tumor markers (HCG and AFP) should be obtained in all patients with germ cell tumors of the testis. These tumor markers have increased the accuracy of diagnosis and clinical staging and have substantially improved our ability to monitor the course of patients with advanced disease. Nevertheless, long experience with the use of tumor markers has demonstrated their limitations as well.

Serum markers should be obtained before orchiectomy whenever possible. Although marker levels cannot be relied upon to rule out malignancy in a patient with a testicular mass, elevated levels may hasten the decision for orchiectomy. Markers should be obtained before any other therapy and at periodic intervals thereafter.

Although the markers may be elevated in some patients with metastases not otherwise clinically detectable, the sensitivity of current assays is limited (20, 23). Table 2 depicts a disturbing 50% false negative rate in pathologic Stage B_1 and 36% in Stage B_2 (23). Hence, marker levels, even when combined with other clinical staging studies, are not sufficiently accurate to obviate surgical staging. Patients with advanced disease almost invariably have elevated tumor markers initially: 94% in our experience (Table 2) (23). Fortunately, no proven false positive values for HCG and AFP have been reported in patients free of tumor, although posttransfusion hepatitis (AFP) and postorchiectomy elevation of pituitary gonadotropin (HCG) may be misleading. A detailed analysis of marker values before treatment revealed that neither the presence nor the absolute value of the marker was of prognostic significance when the impact of stage was considered.

The most effective use of the markers is to monitor the response of the tumor to therapy and to detect early recurrence. Marker levels should be determined before and after each therapeutic intervention and at 2–3 month intervals in patients at high risk of recurrence. In monitoring serum markers

Table 2 Frequency of elevated levels in serum samples obtained after orchiectomy, but immediately before any other therapy, in 142 patients with nonseminomatous germ cell tumors (from Ref. 20)

Stage	Patients (no.)	HCG (%)	AFP (%)	Either (%)
A	67	7	9	10
B_1	15	29	33	50
B_2	23	52	43	64
B_3	6	83	75	100
C	31	84	60	93

it must be remembered that the time required for an elevated marker to return to normal depends upon its serum half-life and its level before treatment. The half-life of HCG is approximately one day and of AFP 3–5 days (12, 19). High pretreatment levels of AFP may take several weeks to return to normal. If the rate of fall of a marker after surgical resection of the tumor is too slow, residual tumor is present. If the marker levels plateau or rise, persistent or recurrent disease may be found. Immunocytochemical studies have demonstrated the presence of HCG only in syncytiotrophoblastic giant cells and of AFP only in embryonal carcinoma or yolk sac tumor; hence elevated marker levels indicate the presence of malignant elements (11).

Nevertheless, the markers are of limited value in the majority of patients initially seen for therapy (23). False negatives diminish their usefulness in clinical staging, and they have no prognostic significance. A return of elevated markers to normal is not necessarily associated with a disappearance of all tumor. Nearly 80% of our patients in whom marker values returned to normal after chemotherapy still had histological evidence of malignant tumor, usually embryonal carcinoma, present at the time of surgical resection (23).

Clinical Management

Patients with nonseminomatous tumor are designated as having either early or advanced disease as indicated above. Those with early stage disease are treated with immediate retroperitoneal lymphadenectomy, which establishes the pathologic stage, removes the nodal metastases, eliminates local retroperitoneal recurrence, and obviates the need for postoperative radiotherapy. Figure 2 depicts the treatment scheme for patients with clinical Stage A and B disease.

It is our practice to follow patients every two months for the first year after lymphadenectomy, every three months for the second year, and then

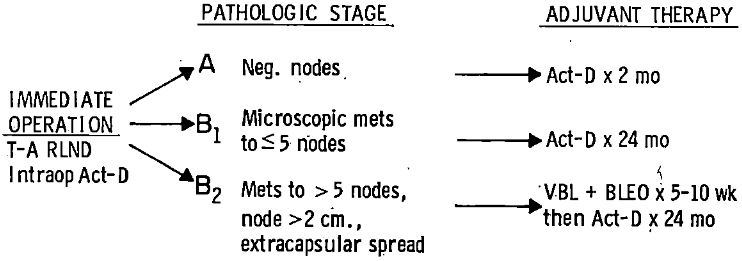

PATHOLOGIC STAGE
ADJUVANT THERAPY
IMMEDIATE OPERATION
T-A RLND
Intraop Act-D
A
Neg. nodes
Act-D x 2 mo
B1
Microscopic mets to ≤ 5 nodes
Act-D x 24 mo
B2
Mets to > 5 nodes, node >2 cm., extracapsular spread
VBL + BLEO x 5-10 wk then Act-D x 24 mo

Figure 2 Management of clinical Stages A and B nonseminomatous germinal tumors of testis; January 1974 to present.

every six months until they are five years post-lymphadenectomy. Follow-up consists of physical examination, a chest x ray, and a serum sample for HCG and AFP. We continue to treat all patients with prophylactic chemotherapy (Figure 2). In this scheme, tumor has persisted or recurred in only 7 of 76 Stage A and B patients treated since 1974 with a minimum follow-up of 36 months. In only three of these seven patients did elevated marker levels actually precede clinical evidence of recurrence. Hence, tumor markers do not replace the need for close clinical follow-up and, perhaps, periodic sampling of tumor markers at subsequent intervals would be more cost effective in these patients—before and after retroperitoneal dissection and, in patients whose markers are negative, again at four months and at one year postoperatively. Serum measurement of HCG is of considerable value, however, in evaluating the etiology of gynecomastia, which has been seen on several occasions in our patients during and following chemotherapy and is usually not due to tumor recurrence. Further experience will be necessary to determine the value of markers in the late postoperative period once prophylactic chemotherapy has been stopped. However, to date, we have not seen late recurrence more than 13 months after initial therapy in patients with initial Stage B or C tumor who remained in sustained complete remission for that period of time.

Pathologic staging (placing patients in groups A, B_1, B_2, B_3, and C) allows selection of a specific chemotherapeutic approach that has achieved a high cure rate associated with minimal morbidity in terms of hospitalization or lost time from work (Figure 2). Intraoperative use of actinomycin D followed by a prophylactic course two months later has resulted in the lowest recurrence rate yet reported for Stage A patients (3.5%) (23). Patients with minimal retroperitoneal disease have done well with actinomycin D as a single agent, with only one death resulting from tumor. More aggressive prophylactic chemotherapy for patients with Stage B_2 disease has reduced the recurrence rate to 12%, significantly lower than that observed when actinomycin D is used alone, or when no adjuvant chemotherapy is given (Table 3). Furthermore, use of vinblastine sulfate and bleomycin followed by cyclic actinomycin D has not decreased possible salvage by the platinum, vinblastine sulfate, and bleomycin combination since only one of 17 patients with Stage B_2 disease has died of metastases. Stage B_3 represents a unique group of patients who enjoy surprisingly high survival provided they are recognized and treated early by intensive chemotherapy (Figure 3) followed by surgical resection. The histology found at the time of a cytoreductive operation has been a reliable indicator of the patients who will require more intensive chemotherapy (platinum, vinblastine sulfate, and bleomycin) compared with those who may be successfully treated with vinblastine sulfate and bleomycin alone.

Table 3 Comparison of fate and survival of 296 patients with pathologic stages A and B_1 to B_2 nonseminomatous germ cell tumors of the testis according to adjuvant therapy (minimum follow-up 36 months) (from Ref. 23)

Adjuvant	Reference	Patients (No.)	Remained same (%)	Recurred stage C		Complete response		Retro-peritoneal recurrence (No.)	Complete response, free of tumor	
				(No.)	(%)	(No.)	(%)		(No.)	(%)
				Stage A						
No chemotherapy and no radiation	26	45	80	9/45	20	6/9		0	42/45	93
	3	30	90	3/30	10	3/3		0	30/30	100
	26	26	77	6/26	23	2/6		0	22/26	85
	Totals	101	82	18/101	18	11/18	61	0	94/101	94
Actinomycin D and no radiation	26	22	100	0/22	0	—		0	21/22[a]	95
	23	56	96	2/56	4	2/2		0	56/56	100
	Totals	78	97	2/78	3	2/2	100	0	77/78[a]	99
				Stages $B_1 + B_2$						
± Actinomycin D	26 (surgery only)	20	75	5/20	25	0/5		2	15/20	75
± Radiation	3	25	56	11/25	44	6/11		1	20/25	80
	23	28	71	8/28	29	2/8		2	22/28	79
	Totals	73	67	24/73	33	8/24	33	5	57/73	78
Aggressive prophylactic chemotherapy	26	16	88	2/16	12	1/2		0	15/16[a]	94
	23	28	86	4/28	14	3/4		0	26/28	93
	Totals	44	87	6/44	13	4/6	66	0	41/44[a]	93

[a] One death from other cause, not tumor.

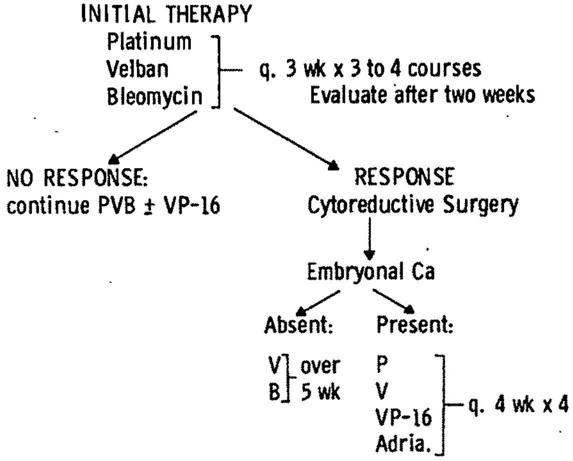


Figure 3 Current management protocol for clinical Stage B_3 or C nonseminomatous germinal tumors of testis.

Therefore, lymphadenectomy with surgical staging remains prominent in patient management since it determines the appropriate use of adjuvant chemotherapy. Furthermore, a meticulous node dissection has been highly effective in eliminating local recurrence. None of the 35 patients with pathologic Stage B disease treated since 1974 has had retroperitoneal recurrence and none has received postoperative radiation therapy.

Current controversy continues in the area of adjuvant therapy of early stage disease. Because patients whose tumors recur after lymphadenectomy are still curable with intensive combination chemotherapy (platinum, vinblastine sulfate, and bleomycin), some urologic oncologists and medical oncologists recommend only close observation following lymphadenectomy (3, 4). There is no question that the combination of platinum, vinblastine sulfate, and bleomycin is currently the most effective available in the management of nonseminomatous germinal tumors of the testis and should be used early in the patient with massive disease (Figure 3). It also is the most toxic regimen, with substantial early morbidity, occasional mortality, and uncertain long-term effects. *Cis*-platinum may cause permanent renal impairment and is a mitogen that induces long-term bone marrow suppression; it is unknown whether induction of malignancy seen following use of alkylating agents will be observed in patients receiving *cis*-platinum (10). Our results indicate that excellent control of these tumors is possible without platinum in the great majority of patients, and that actinomycin D is probably a valuable agent for patients with minimal disease. Prophylactic adjuvant chemotherapy can lower the recurrence rate significantly without apparent long-term morbidity. The early aggressive use of the combination of vinblastine sulfate and bleomycin remains highly effective in preventing pulmonary metastases in patients with extensive abdominal disease and has resulted in a recurrence rate of 12% for Stage B_2 disease. Table 4 lists the fate of 296 patients with pathologic Stages A and B_1 or B_2 disease treated at three centers by meticulous lymphadenectomy. Note that the use of adjuvant chemotherapy reduced the recurrence rate in Stage A from 18 to 3%, and that early aggressive prophylactic chemotherapy reduced the recurrence rate in Stages B_1 and B_2 from 33 to 14%. Therefore, most patients can be spared the need for platinum, yet those who fail can still be salvaged. In our opinion, the recurrence rate of 44% reported by Donohue and associates (3) in patients with Stages B_1 and B_2 disease is too high to justify withholding early adjuvant treatment, even though most recurrences can be controlled subsequently by platinum in combination.

While the future may offer a more limited node-sampling procedure in an effort to preserve fertility, current evidence suggests that it is the patients with the most limited microscopic disease that benefit most from meticulous surgical staging and may, thereby, avoid recurrence and the need for intensive platinum, vinblastine sulfate, and bleomycin chemotherapy. This point should be emphasized in lieu of the lack of sensitivity of biochemical tumor markers for minimal disease and our inability to carefully assess and monitor the retroperitoneal nodes. A meticulous surgical staging procedure currently allows selection of the least toxic chemotherapeutic agent or combination tailored to maximum survival of the individual patient, eliminates retroperitoneal recurrence, and remains a most important factor in dictating therapeutic plans.

MANAGEMENT OF ADVANCED DISEASE

The development of truly effective combination chemotherapy has remarkably changed the expected survival of patients with advanced disease. Prior to 1973 the reported cure rate for patients with B_3 or C disease was less than

Table 4 Tumor-free survival of 227 consecutive patients with pathologic Stage A and B nonseminomatous germ cell tumors treated since 1974 at three centers by meticulous retroperitoneal lymphadenectomy and chemotherapy without radiation therapy

	Stage A		Stage B		Total A + B		Retroperitoneal
Series	(No.)	(%)	(No.)	(%)	(No.)	(%)	recurrence
Staubitz (26)	21/22	95	15/16	94	36/38	95	0
Donohue et al (3)	57/57	100	54/55	98	111/112	99	1
Skinner & Scardino (23)	42/42	100	32/35[a]	91	74/77	96	0
Total	120/121	99	101/106[a]	95	221/227	97	1

[a] One death from other cause, not tumor. There are no exclusions due to extensive retroperitoneal disease. Minimum follow-up 3 years.

10%; today more than 70% can be expected to be cured of their disease. The role of surgery in the management of advanced disease is less well defined and remains somewhat controversial. Without question, the combination of platinum, vinblastine sulfate, and bleomycin has been so impressive in achieving a complete response rate that surgery has become adjunctive to the primary curative goals of chemotherapy. Patients with widely disseminated disease, with or without bulky retroperitoneal metastases, should be treated initially and vigorously by intensive combination chemotherapy, utilizing the Einhorn & Donohue (4) protocol modified by reduction in the dosage of vinblastine sulfate to 0.3 mg/kg for each course of therapy, and further attenuated for low performance status (4, 10). Nonetheless, surgery plays an important role in the management of these patients by removing all retroperitoneal disease, and providing histologic material to better assess the effects of chemotherapy, occasionally indicating the need to alter drugs.

Experience at UCLA in the treatment of advanced disease by the combination of platinum, vinblastine sulfate, and bleomycin alone has not been as good as that reported by Einhorn, with most failures occurring in those patients harboring any pulmonary metastases greater than 2.5 cm in diameter or initially having bulk retroperitoneal abdominal disease (18). Some authors advocate cytoreductive surgery before chemotherapy with the hope of diminishing tumor burden and thereby rendering chemotherapy more effective (13). Early trials at initial massive debulking resulted in significant complications in terms of tumor spill, prolonged postoperative recovery, and the catabolism of these postoperative patients that delayed initiation of needed effective chemotherapy. Furthermore, retroperitoneal resection of large tumor masses does not lend itself to early control of a vascular pedicle as in other tumors, thereby increasing the risks of tumor dissemination with surgical manipulation. It seemed appropriate, therefore, to treat such patients preoperatively with chemotherapy. Beginning in the early 1970s, our utilization preoperatively of older, less effective drugs indicated that combination chemotherapy could significantly shrink the tumor bulk and produce a thick fibrous capsule or desmoplastic reaction around the tumor mass, thereby making resection possible in all cases without either tumor spill or leaving any residual tumor. In fact, in our 35 patients with massive Stage B_3 and C disease, under a protocol that did not include *cis*-platinum, 29 (83%) achieved a complete response and 26 (74%) of these patients remain free of tumor with follow-up from 30 to 102 months (20). Since 1977 all patients have been managed by the protocol illustrated in Figure 3 utilizing 3–4 courses of preoperative platinum, vinblastine sulfate, and bleomycin. Surgery is performed 2–3 weeks following the fourth course of therapy in all responding patients through a thoracoabdominal approach described previously, with wedge resection of persistent ipsilateral pulmonary nodules performed at the same time. If pulmonary disease persists in the contralateral lung and is amenable to surgical resection, a formal contralateral thoracotomy is then performed 4–6 weeks later. Median sternotomy is an alternative approach that allows simultanous resection of bilateral disease as well as resection of any residual mediastinal disease. Histology of the resected tissue reveals persistent embryonal carcinoma in approximately one-third of these patients, mature teratoma with or without extensive fibrosis and necrosis in approximately one third, and fibrosis only without any evidence of tumor in approximately one third. It has been impossible to predict on the basis of CT scanning, preoperative serum tumor markers, or careful inspection at the time of surgery, who harbors embryonal carcinoma vis à vis teratoma or fibrosis. Obviously those patients with elevated markers will have embryonal carcinoma, but in most patients the markers fall to normal during chemotherapy and a substantial number of patients with normal markers still have evidence of embryonal carcinoma at operation.

It continues to be our policy to operate on all patients with advanced disease following 3–4 courses of platinum, vinblastine sulfate, and bleomycin protocol despite clinical evidence of a complete response by all noninvasive parameters. Others recommend surgery for only those patients with an abnormal CT scan or other evidence of persistent disease. Our reason for recommending surgery for all patients is twofold: first, to assure that all retroperitoneal disease is controlled, thus preventing late failure in an area that is extremely difficult to monitor; second, the tissue obtained at surgery can be examined histologically and utilized in the in vitro clonogenic assay described by Salmon to further tailor effective combination chemotherapy and to dictate its subsequent need as well as the most appropriate selection of drugs. For example, based on histologic examination of this tissue, we have found that those patients with only mature teratoma or fibrosis without evidence of embryonal carcinoma do well without the need for continued intensive combination chemotherapy utilizing platinum. On the other hand, those patients with persistent embryonal elements, or other immature elements, clearly need further chemotherapy since fewer than 50% of these patients survive despite additional courses of platinum, vinblastine sulfate, and bleomycin. The in vitro clonogenic assay is being routinely used in these patients to help select subsequent chemotherapy. In this way it may be possible to develop novel combinations individually tailored to the drug sensitivity profiles of the residual tumor. A wide variety of additional agents including Mithramycin, doxorubicin hydrochloride, VP-16-213 may be identified as appropriate.

Another consideration is the extent of cytoreductive surgery following chemotherapy. Is biopsy only or "lumpectomy" sufficient, or should the operation be a meticulous dissection similar to that for Stage A and B disease? Our experience and that of others indicate that the dense desmoplastic reaction around the great vessels and residual tumor bulk, as well as the cystic changes that occur following chemotherapy, make it impossible to judge by gross inspection just what tissue harbors persistent tumor. Foci of embryonal carcinoma have been found in association with solid fibrosis, cystic, or necrotic tissue. Therefore, there seems no feasible method to identify physically or by palpation what in actuality remains a microscopic diagnosis. Therefore, we recommend as complete and meticulous a retroperitoneal lymph node dissection as feasible in each case. The results of an integrated management plan currently yield a complete response rate in excess of 80% for all patients who present with massive abdominal disease.

CONCLUSIONS

In summary, the 1970s saw dramatic changes in the management of nonseminomatous germ cell tumors of the testis resulting in a tumor-free survival of 97% for all patients with Stage A and B disease, and a tumor-free survival in excess of 70% for patients with advanced disseminated disease. Biochemical tumor markers are helpful in the management of advanced disease and in evaluating clinical suspicion of tumor recurrence, but they have a more limited value in patients who initially present with early stages (A and B) of nonseminomatous germ cell tumors of the testis, since their influence on the therapeutic regimen is minimal and the treatment failure rate in optimally managed patients is so low.

Improved chemotherapy utilizing platinum, vinblastine sulfate, and bleomycin is largely responsible for the remarkable improvement in survival, but a meticulous lymphadenectomy remains the most important determinant in selecting the least toxic chemotherapeutic adjuvant, eliminating local retroperitoneal recurrence, and obviating the need for radiation therapy.

Literature Cited

1. Deleted in proof
2. Caldwell, W. Z., Kademian, M. T., Frias, Z., Davis, T. E. 1980. The management of testicular seminomas. *Cancer* 45:1768–74
3. Donohue, J. P., Einhorn, L. H., Perez, I. M. 1978. Improved management of nonseminomatous testis tumors. *Cancer* 42:2903–8
4. Einhorn, L. H., Donohue, J. P. 1979. Combination chemotherapy in disseminated testicular cancer: the Indiana University experience. *Semin. Oncol.* 6:37–93
5. Friedman, N. B. 1977. Pathology of testicular tumors. *West. J. Med.* 126:362
6. Friedman, N. B., Moore, R. A., 1946. Tumors of the testis: a report on 922 cases. *Mil. Surg.* 99:573
7. Deleted in proof
8. Henderson, B. E., Benton, B., Jing, J., Yu, M. C., Pike, M. C. 1979. Risk factors for cancer of the testis in young men. *Int. J. Cancer* 23:598–602

9. Kaswick, J., Bloomberg, S., Skinner, D. G. 1976. Retroperitoneal lymph node dissection: how effective in removal of all retroperitoneal nodes? *J. Urol.* 115:70
10. Krikorian, J. G., Daniels, J. R., Brown, B. W. Jr., et al. 1978. Variables for predicting serious toxicity (vinblastine dose, performance status, prior therapeutic experience): chemotherapy for metastatic testicular cancer with *cis*-dichlorodiammineplatinum (II), vinblastine, and bleomycin. *Cancer Treat. Rep.* 62:1455–63
11. Kurman, R. J., Scardino, P. T. 1980. Immunoperoxidase localization of alpha fetoprotein and human chorionic gonadotropin in malignant germ cell tumors of the ovary and testis. In *Diagnostic Immunocytochemistry,* ed. R. Delallis. New York: Masson
12. Lange, P. H., McIntire, K. R., Waldmann, T. A., et al. 1976. Serum alpha fetoprotein and human chorionic gonadotropin in the diagnosis and management of nonseminomatous germ cell testicular cancer. *N. Engl. J. Med.* 195:1237–40
13. Merrin, C., Takita, H., Beckley, S., Kassis, J. 1977. Treatment of recurrent and widespread testicular tumors by radical reductive surgery and multiple sequential chemotherapy. *J. Urol.* 117:291
14. Deleted in proof
15. Mostofi, F. K. 1980. Pathology of germ cell tumors of testis: a progress report. *Cancer* 45:1735–54
16. Samuels, M. L., Johnson, D. E., Holoye, P. Y. 1973. The treatment of Stage C metastatic germ cell neoplasia of the testis with bleomycin combination chemotherapy. *Proc. Am. Assoc. Cancer Res.* 14:23
17. Deleted in proof
18. Sarna, G., Skinner, D. G., Smith, R. B., Zighelboim, J., Goodnight, J. E., Feig, S. 1980. *Cis*-platinum diammine dichloride alone and in combination in testicular and other malignancies. *Cancer Treat. Rep.* In press
19. Scardino, P. T., Cox, H. D., Waldmann, T. A., et al. 1977. The value of serum tumor markers in the staging and prognosis of germ cell tumors of the testis. *J. Urol.* 118:994–99
20. Scardino, P. T., Skinner, D. G. 1979. Germ cell tumors of the testis: improved results in a prospective study using combined modality therapy and biochemical tumor markers. *Surgery* 86:86–93
21. Shipley, W. U. 1980. Role of radiation therapy in the management of adult germinal testis tumors. In *Testicular Tumors, Management and Treatment,* ed. L. H. Einhorn, pp. 47–67. New York: Masson
22. Skinner, D. G. 1978. Management of nonseminomatous tumors of the testis. In *Genitourinary Cancer,* ed. D. G. Skinner, J. B. deKernion, p. 470. Philadelphia: Saunders
23. Skinner, D. G., Scardino, P. T. 1980. Relevance of biochemical tumor markers and lymphadenectomy in the management of nonseminomatous testis tumors: current perspective. *J. Urol.* 123:378–82
24. Smith, R. B. 1978. Diagnosis and staging of testicular tumors. See Ref. 22, p. 448
25. Smith, R. B. 1978. Management of testicular seminoma. See Ref. 22, p. 460
26. Staubitz, W. 1979. Personal communication. Discussed in Ref. 23

Ann. Rev. Med. 1981. 32:559–81
Copyright © 1981 by Annual Reviews Inc. All rights reserved

CEPHALOSPORINS

◆7419

Barbara E. Murray, M.D.

Program in Infectious Diseases and Clinical Microbiology, The University of
Texas Medical School, Houston, Texas 77025

Robert C. Moellering, Jr., M.D.

Infectious Disease Unit, Massachusetts General Hospital, Harvard Medical
School, Boston, Massachusetts 02114

Introduction

There are currently over a dozen cephalosporins and cephalosporin-like antibiotics commercially available throughout the world. Although the first 20 years of research on these agents produced compounds of rather similar activity, the past decade, and in particular the past five years, has seen the development of a host of different semisynthetic cephalosporin and cephamycin derivatives with widely varying pharmacologic and in vitro activities. The choices currently facing the practicing physician are numerous and will almost surely become greater over the next few years. We discuss here the common features of this antimicrobial family and review the individual characteristics and advantages of the currently available drugs, with a brief preview of the more promising agents of the future.

Historical Background

In 1945 Giuseppe Brotzu isolated a fungus, later identified as *Cephalosporium acremonium,* from the Mediterranean Sea off the isle of Sardinia (1, 2). This fungus elaborated antibacterial compounds, including cephalosporin P and cephalosporin (or penicillin) N, but these were not sufficiently active to warrent development into clinically useful antibiotics. Further work by Abraham and Newton with substances produced by *C. acremonium* led to the discovery of cephalosporin C, which is resistant to staphylococcal β-lactamase and inhibits the growth of many gram-positive and gram-negative bacteria. Most of the currently available cephalosporin-like compounds have resulted from chemical modification of the cephalosporin C nucleus. An important exception is cefoxitin, a semisynthetic

0066-4219/81/0401-0559$01.00

derivative of cephamycin C that is derived from the fungus *Streptomyces lactamdurans.* In addition, among the new drugs currently undergoing clinical investigation is a new, totally synthetic drug (moxalactam), which is a 1-oxacephem compound closely related to the cephalosporins.

Classification

The true cephalosporins, excluding cefoxitin, a cephamycin, all share a common nucleus, 7-aminocephalosporanic acid (Figure 1). This consists of a four-member β-lactam ring common also to the penicillins, and a six-member sulfur-containing dihydrothiazine ring. Chemical modifications at the 3 and 7 positions of the 7-aminocephalosporanic acid nucleus produce cephalosporins that vary widely in their spectrum of activity and pharmacokinetic properties. The basic nucleus of cefoxitin, a derivative of cephamycin C, differs slightly from 7-aminocephalosporanic acid: it has a 7-α-methoxy group, which markedly enhances the stability of the β-lactam ring to hydrolysis.

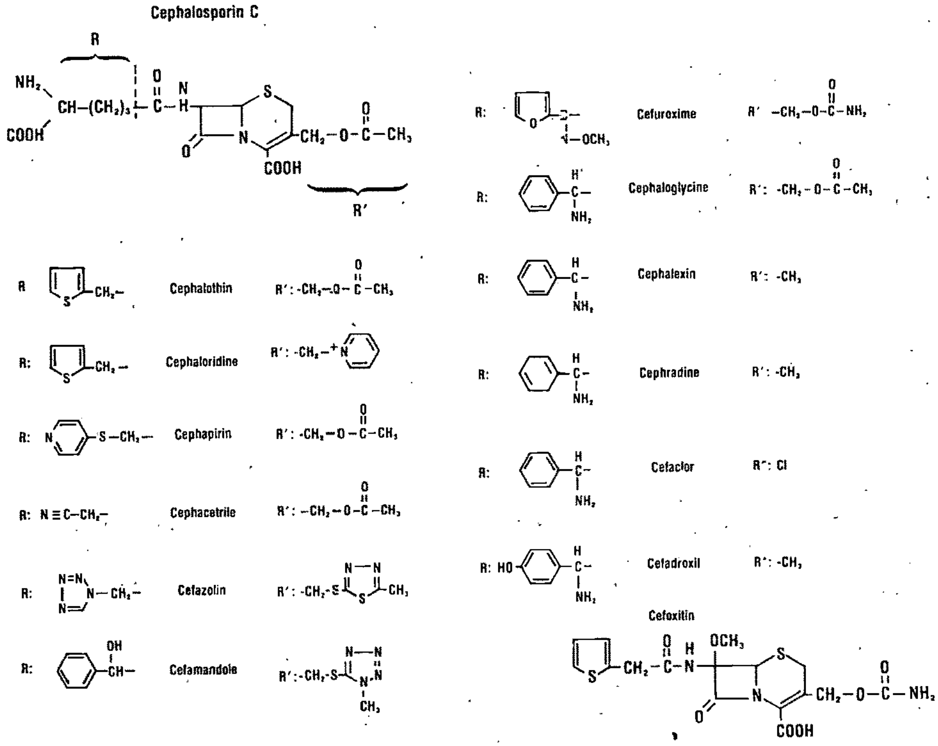


Figure 1 The structure of the true cephalosporins.

The currently available cephalosporin and cephamycin antibiotics are listed in Table 1. The "first-generation" cephalosporins include cephalothin, cephaloridine, cephaloglycine, cephalexin, cephradine, cefazolin, cephapirin, cephacetrile, and cefadroxil. These compounds all share similar spectra of activity and differ primarily in their pharmacokinetic properties. Newer or second-generation cephalosporin-like compounds with an expanded spectrum of activity include cefamandole, cefoxitin, and cefaclor, which are available in this country, and cefuroxime, which is available in Europe and elsewhere. The third-generation cephalosporins, still under investigation in the US, are discussed in a later section.

Mechanism of Action

Although the mechanism of action of penicillin has been studied more extensively than that of cephalosporins, these β-lactam compounds appear to inhibit bacteria in a similar manner. Until recently the key features of the effects of β-lactams were thought to be as follows: (*a*) β-lactams structurally resemble D-alanyl-D-alanine, a unique component of the prokaryotic cell wall and a substrate of the enzyme transpeptidase; (*b*) inhibition of transpeptidase by these structural analogs prevents cross-linking of new murein subunits; (*c*) absence of cross-linkage results in a structurally weakened cell wall; (*d*) a weakened cell wall permits the accumulation of cytoplasmic mass and the subsequent rupture of the cell wall due to increased osmotic pressure. Other morphologic changes caused by penicillin, such as bulge formation, formation of round cells and of filaments, were considered

Table 1 Cephalosporins (and cephamycin)—nomenclature and routes of administration

Generic name	Trade name	Route of administration[a]
Cefamandole	Mandol	IM, IV
Cefazolin	Ancef; Kefzol	IM, IV
Cefoxitin	Mefoxin	IM, IV
Cefuroxime	—[b]	IM, IV
Cephacetrile	Celospor[b]	IM, IV
Cephalexin	Keflex	PO
Cephaloglycine	Kafocin	PO
Cephaloridine	Loridine	IM, IV
Cephalothin	Keflin	IM, IV
Cephapirin	Cefadyl	IM, IV
Cephradine	Anspor, Velosef	PO, IM, IV
Cefaclor	Ceclor	PO
Cefadroxil	Duricef	PO

[a] Abbreviations: PO = per os; IM = intramuscular; IV = intravenous.
[b] Not available in US.

trivial consequences of the structurally defective wall (3). Over the past decade, detailed research on many aspects of the action of penicillins has led to a much more complex and not yet completely understood view. It is known now, for example, that there are several enzymes whose functions are inhibited by penicillin (carboxypeptidases, endopeptidases, and transpeptidases); the number of these appears to vary among species. In addition, there are penicillin binding proteins (PBP), some of which appear to be forms of the above enzymes; these also vary among species (*S. aureus* has four, *Bacillus subtilis* has five, and *E. coli* has six) (4). Different β-lactams bind preferentially to different PBPs and this appears to correlate with specific physiologic effects. Benzyl penicillin, for example, binds preferentially to PBP3 at low concentrations and this is associated with selective inhibition of cell division. Cephaloridine binds PBP1 at low concentrations and causes rapid cell lysis. Mecillinam binds preferentially to PBP2 and causes the formation of ovoid forms. Spratt (5) created temperature-sensitive mutants of the PBP genes and can cause penicillin-inducible morphologic changes to appear and disappear by varying the temperature of incubation, again supporting the role of PBPs.

Besides causing bacteriostasis (i.e. inhibition of bacterial cell growth), β-lactam antibiotics typically cause loss of viability of intact cells, and in many species will cause complete cellular lysis. Tomasz (3, 6) proposed that the eventual lysis of bacterial cells caused by β-lactams and other cell wall active agents results because these antibiotics allow the cells to release inhibitors of endogenous autolysins (murein hydrolases); this then allows the unopposed cleavage of the cell wall murein with subsequent lysis (or autolysis) of the cell. Evidence for this theory comes from studies with pneumococci; mutants of pneumococci have been isolated that do not lyse after exposure to penicillin but that otherwise have the same inhibitory and morphologic responses to penicillin. These mutants have suppressed N-acetylmuramyl-L-alanine amidase activity; this enzyme is the major murein hydrolase (autolysin) of pneumococci, and absence of its activity appears to cause defective lysis. If the mutant cells are exposed to wild-type murein hydrolase plus penicillin, then lysis will occur. Furthermore, if one takes pneumococcal murein hydrolase inhibitor and coats wild-type cells with an excess of this inhibitor, then subsequent penicillin-induced lysis is blocked. Just how murein hydrolase inhibitors function and how β lactams labilize and release these inhibitors from the bacterial cell is not yet known.

Mechanism of Resistance

In order for an antimicrobial agent to inhibit bacterial growth, it must be able to reach a susceptible target site(s) in sufficient quantities. Bacteria, therefore, may resist the effects of antimicrobial agents by preventing their uptake, by inactivating them before they can act upon the target site(s), or by not having a target site that is susceptible to the effects of the antibiotic (7). β-Lactam antibiotics, for example, must reach the penicillin-binding proteins and the enzymes associated with peptidoglycan (murein layer) formation. Gram-negative bacteria have an outer phospholipid membrane surrounding the murein layer that can retard the entry of various substances, including β-lactam antibiotics (8). Penicillin G has more difficulty penetrating the outer membrane of gram-negative bacilli than do ampicillin and the cephalosporins, and most gram-negatives are therefore relatively resistant to penicillin G. Alteration of the β-lactam target site(s) (the PBPs and the peptidoglycan enzymes) or variations in the affinity of various β-lactams for the target site(s) also lead to varied morphologic responses, and even resistance to these antibiotics (3, 9).

The major mechanism by which bacteria resist β-lactam antibiotics is the production of β-lactamase(s), a group of enzymes that can hydrolyze the β-lactam ring and render the antibiotic inactive. The effectiveness of a β-lactamase in conferring resistance upon the producing strains depends on an interaction of various factors, including affinity of the enzyme for the β-lactam antibiotics, velocity of the hydrolysis reaction, and, probably most importantly, the rate of delivery of the antibiotic to the enzyme, which in turn seems to be governed by the outer phospholipid membrane present in gram-negative organisms (7). Consequently, β-lactams that are efficiently hydrolyzed in vitro by a β-lactamase may still be effective in vivo against strains producing that enzyme if there is no barrier function limiting uptake of the β-lactam; the enzyme system can become saturated or overwhelmed by the quantity of antibiotic present. Conversely, a strain with an effective permeability barrier but a relatively ineffective β-lactamase may be resistant because the rate of delivery of antibiotic is sufficiently slow to allow complete inactivation (10).

The β-lactamases produced by gram-positive organisms (*S. aureus* being the most widely studied and the most important clinically) are primarily penicillinases, active against penicillin and ampicillin but not against the penicillinase-resistant semisynthetic penicillins (PRSP) such as methicillin, nafcillin, and oxacillin, and not against most cephalosporins. Most staphylococci therefore are susceptible to the cephalosporins; methicillin-resistant *S. aureus* are an exception, but this resistance is not due to penicillinase production (11–13).

Some cephalosporins, notably cephaloridine and cefazolin, are known to be more susceptible to hydrolysis by *S. aureus* β-lactamase than other cephalosporins. Cephalothin is the most resistant to hydrolysis by these enzymes with cefoxitin and cefamandole being next most resistant. The clinical relevance of these observations is not clear although there is some evidence to suggest that the more susceptible cephalosporins may be less effective for infections caused by β-lactamase-producing *S. aureus* (14). Until more information is available, if a cephalosporin is indicated for an infection due to β-lactamase-producing staphylococci, it is probably most prudent to use an agent such as cephalothin that is highly resistant to staphylococcal β-lactamase and is of well-established efficacy (2).

Gram-negative bacilli produce a vast array of β-lactamases. Some are predominantly active against penicillins while others are very active against cephalosporins; some are coded for by plasmid genes and others by chromosomal genes. Of the currently available agents discussed in this paper, cefoxitin, a cephamycin, is the most resistant to hydrolysis by these β-lactamases. Using the classification of Richmond & Sykes (14a), cefoxitin is usually resistant to hydrolysis by types I, II, III (including TEM β-lactamase, a common broad spectrum enzyme type), IV, and V, although occasional enzymes are found that can hydrolyze cefoxitin. Cefuroxime (available in some parts of the world) is the second most resistant to β-lactamase hydrolysis. Cefamandole is more resistant than cephalothin but less resistant than cefoxitin. The oral agent, cefaclor is also resistant to hydrolysis by TEM (type IIIa) β-lactamase, the enzyme produced by β-lactamase-producing *H. influenzae* and *N. gonorrheae* but is hydrolyzed by type I chromosomal cephalosporinases (2, 15–21). As mentioned above, the in vitro activity of partially purified β-lactamase from a given organism fails to show 100% correlation with the results of in vitro susceptibility testing for that strain since susceptibility results from a combination of permeability, enzymatic modification, and susceptibility of target site.

Spectrum of Activity

Cephalosporins and cephamycins are truly broad spectrum antibiotics; as such, they are rarely the drug of choice for therapy of an established infection caused by a single organism. Most staphylococci, including β-lactamase-producing strains, are inhibited by the first-generation cephalosporins. Methicillin-resistant *S. aureus* are resistant to cephalosporins; results with methicillin-resistant *S. epidermidis* vary (22). Cephalosporins are active in vitro against almost all streptococci except enterococci and penicillin-resistant pneumococci. The majority of strains of *E. coli, K. pneumoniae,* and *P. mirabilis,* especially those that are community acquired, are susceptible to cephalosporins, whereas most of the remaining aerobic and facultative gram-negative bacilli, as well as some hospital-acquired R-factor-containing strains of *E. coli, K. pneumoniae,* and *P. mirabilis,* are resistant. Most anaerobic species are susceptible in vitro except for *Bacteroides fragilis* (2, 23).

The currently available parenteral second-generation cephalosporin/cephamycin antibiotics (cefamandole, cefoxitin, and cefuroxime) have expanded activity against gram-negative organisms and can be active against many cephalothin-resistant *E. coli, Klebsiella,* and *P. mirabilis,* against *H. influenzae* and *N. gonorrheae,* including β-lactamase-producing strains, and against variable percentages of *Citrobacter freundii, Enterobacter,* indole-positive *Proteus* sp., *Providencia stuartii,* and some *Serratia* sp. (23–29). Cefoxitin also has enhanced activity against *B. fragilis* (30, 31). None of these antibiotics is active against *Pseudomonas aeruginosa* and none is sufficiently active against strains of *Acinetobacter* sp., *Serratia* sp, and other multiply resistant (frequently hospital-acquired) gram-negative bacilli; some of the new cephalosporins under investigation (discussed below) are of particular interest because of their enhanced activity against pseudomonas and serratia.

Pharmacology

ROUTE OF ADMINISTRATION Manipulations of the side chains on the β-lactam ring of cephalosporins produce compounds with varying pharmacologic properties. Of the currently available cephalosporins, cephalothin, cephapirin, cefoxitin, cephaloridine (all available in the US), cefuroxime, and cephacetrile (available in Europe) must be administered parenterally. Cephradine may be administered orally or parenterally. Cephalexin, cephradine, cefaclor, and cefadroxil produce adequate systemic levels after oral ingestion; cephaloglycine is also available for oral use but produces adequate levels only in the urine. Pain after injection is a major factor limiting the intramuscular administration of some cephalosporins; in general, cefazolin and cephaloridine are better tolerated than the others by this route (23).

After the administration of one gram intramuscularly, cephalothin, cephapirin, cephacetrile, and cefoxitin will produce peak serum levels of approximately 20 μg/ml (32). Cefamandole produces slightly higher levels of 20–35 μg/ml (33); cefuroxime and cephaloridine give peak values of up to 40 μg/ml; and cefazolin produces peak serum levels of 50–75 μg/ml (2, 23, 32, 33). A summary of major pharmacologic properties is shown in Table 2.

METABOLISM AND EXCRETION Most cephalosporins do not undergo in vivo metabolism and are excreted unchanged in the urine, but some (cephalothin, cephaloglycine, cephacetrile, and cephapirin) have an acetyl group at the 3 (R^1) position and undergo desacetylation in vivo to form desacetyl derivatives that are less active than the parent compounds. Cefoxi-

Table 2 Cephalosporins (and cephamycin)—serum protein binding, peak serum concentrations, and serum half-life (T1/2 β) after intramuscular administration of a one-gram dose or oral administration of a 500-mg dose

Compound	Protein bound (%)	Route of administration	Peak serum level (μg/ml)	Serum T1/2 β (min)	References
Cefaclor	22–25	PO	10–23	30–60	78, 116, 117
Cefamandole	67–80	IM	20–35	55–90	33, 34, 118, 119
Cefazolin	74–86	IM	38–76	96–153	23, 32
Cefoxitin	50–73	IM	20–22.5	41–48	23, 32, 33, 34
Cefuroxime[a]	28–38	IM	34–40	65–82	33, 34, 115, 120
Cephacetrile[a]	33–36	IM	22–22.7	33–90	23, 32
Cephalexin	10–15	PO	12–18	36–54	23, 32
Cephaloglycine	"moderate"	PO	0–2	—	32
Cephaloridine	8–31	IM	38–40	48–90	23, 32
Cephalothin	65–79	IM	15–21	28–51	23, 32, 33, 34, 119
Cephapirin	44–50	IM	15–24	37–47	23, 32
Cephradine	6–20	PO	11–22	31–58	23, 32
Cephradine	6–20	IM	10.4	31–58	32
Cefadroxil	17–20	PO	17.8	80–97	81, 88

[a] Marketed elsewhere, but not available in US. Abbreviations: PO = per os; IM = intramuscular.

tin undergoes negligible metabolism to its descarbamyl derivative (32, 34).

Cephalosporins and their metabolic products are generally excreted into the urine by both glomerular filtration and tubular secretion. The only exception is cephaloridine, which does not undergo tubular secretion. Impaired renal function leads to the accumulation of cephalosporins and potentially to toxic effects, although documented toxicity in these patients is uncommon (35). Recommendations for dose reduction relative to the degree of renal impairment are somewhat varied but general guidelines based on recent reviews and the manufacturer's recommendations are summarized in Table 3 (2, 35). The administration of probenicid blocks tubular secretion of many agents including cephalosporins, which increases serum levels and prolongs the half-life. Since cephaloridine is secreted only negligibly by the renal tubules, its excretion rate is not appreciably altered by probenecid.

ORGAN PENETRATION Cephalosporins appear to penetrate well into pleural, pericardial, and synovial fluid, into most other tissue spaces except CSF, and across the placenta (32, 34). The effect of protein binding of an antibiotic upon its distribution into fluid spaces has been studied by various authors. Although some studies show an inverse correlation between the percentage of protein binding and penetration into inflammatory peritoneal

Table 3 "Maximum" daily maintenance dosages of selected cephalosporins in patients with varying renal function[a]

	Creatinine clearance (ml/min)					
	Normal (>80)	80–50	50–25	25–10	10–2	<2
Cefazolin	4–6[b]	4	1–3	0.5–1.5	0.25–0.75	0.25–0.75
Cefamandole	12	8	6	4	2	1.5
Cefoxitin	12	8	6	4	2	1
Cephaloridine	4	3	1.2–2	0.6	0.4	0.3
Cephalothin	12	8	6	4	2	1.5
Cephapirin	12	4–8	4–8	4–8	2–3	1–2
Cephradine	8	8	4–8	2–4	2	1–2

[a] Maximum doses recommended by manufacturers. In some instances, doses given in the table are approximations calculated from manufacturers recommendations. For exact details see package insert.

[b] Grams per 24 hours.

fluid (36) or into fibrin clots (37), this is not a universal finding. Moreover, an antibiotic that is highly protein bound may in effect be captured within a protein-containing space with little or no back diffusion so that successive doses may result in progressively higher levels (36). This may partially explain the efficacy of cefazolin, a highly protein bound cephalosporin, in achieving high extravascular fluid levels in animal studies (38–40). In a comparative study of the penetration of cephradine (approximately 10% protein bound) and cefazolin (approximately 85% protein bound) into atrial appendage and pericardial fluid, several interesting observations were made. As expected, the total serum concentration of cefazolin exceeded that of cephradine whereas the free serum concentration showed the reverse relationship. The total pericardial fluid concentrations were approximately equal whereas the free pericardial fluid concentrations of cephradine greatly exceeded those of cefazolin. On the other hand, right atrial appendage muscle levels of cefazolin were greater than those of cephradine. The clinical implication of these findings remains obscure since it is not clear whether it is more important to have the higher and more prolonged serum levels achievable with cefazolin, as well as the higher atrial muscle levels observed, or to have the higher free serum and pericardial fluid levels observed with cephradine (41).

Cerebrospinal fluid penetration by cephalosporins is generally poor, even in the face of meningeal inflammation, and bacterial meningitis due to cephalosporin-susceptible organisms has developed in patients receiving cephalothin and cefazolin (23). Cephaloridine gives relatively high CSF concentrations but clinical usage has not established its efficacy (42). Cefamandole also penetrates into the CSF in higher concentrations but clinical results have been mixed (43, 44). Levels of cefoxitin of 4.3–12.3 μg/ml have been achieved in bacterial meningitis when cefoxitin is administered with probenecid (45), and results of therapy from a few studies have been encouraging (46, 47). Cefuroxime levels of $\leq$ 0.09 μg/ml with noninflamed meninges, up to 1.9 μg/ml with moderately inflamed meninges, and up to 3.9 μg/ml with severely inflamed meninges have been observed (48, 49). Until further study, none of these agents should be relied upon for the therapy of bacterial meningitis. Biliary tract penetration by cephalosporins is variable and many of them do not achieve high concentrations. Cefazolin achieves levels that are equal to or higher than its serum levels but cefamandole and cefoxitin produce the highest levels, even though their serum levels are less than those produced by cefazolin (2, 32). Therapeutic levels may be obtained in patients with cholecystitis but are not achieved in the presence of biliary obstruction (50–53).

Adverse Effects

With the exception of a few adverse effects such as dose-related nephrotoxicity, which is specific for cephaloridine, the presently available cephalosporins cause relatively few significant adverse reactions. Local reactions such as pain or phlebitis after intravenous (IV) administration, pain on intramuscular (IM) injection, and nausea or diarrhea after oral administration are relatively common. A number of studies have examined the potential of various cephalosporins and cephalosporin preparations [buffered vs unbuffered solutions (23)] to cause phlebitis after IV administration. The results of these studies have been less than conclusive. There are apparently no major differences among the cephalosporins in the incidence of phlebitis on IV administration, but the phlebitis caused by cephalothin may be more severe (11, 23, 54). Of the presently available compounds, cephaloridine and cefazolin are the best tolerated when given intramuscularly. Cefamandole and cefoxitin may also be given IM but cause more pain than cephaloridine or cefazolin (R. C. Moellering, personal observations). There appear to be no major differences in adverse reactions caused by the various oral cephalosporins.

Of the systemic adverse effects caused by the cephalosporins, hypersensitivity reactions appear to be the most common. Eosinophilia, skin rash, drug fever, serum sickness, and anaphylaxis (the latter very uncommon) have been described in 2–5% of patients receiving cephalosporins (2, 23, 55). If one excludes eosinophilia, the incidence of reactions given above would be cut approximately in half (56). Of practical concern is the incidence of allergic reactions to the cephalosporins in patients with a past history of penicillin allergy. Patients with documented penicillin allergy develop serum antibodies that cross-react with the cephalosporins (55). Most of these are directed against the "major antigenic determinants" (penicilloyl-protein and cephaloyl-protein). Since cephaloyl groups are very labile, they may not remain in serum long enough to play a major role in cross-sensitivity reactions. "Minor antigenic determinants" may be more important in IgE-mediated immediate reactions to the penicillins and cephalosporins (57), but the exact significance of this in determining cross-sensitivity remains to be determined (58). A number of studies suggest that the overall risk of allergic reations to the cephalosporins in patients with documented penicillin allergy is 5–16% (55, 59). However, it is not clear whether this represents a degree of cross-sensitivity or whether it is simply because one sees a higher incidence of hypersensitivity reactions even to immunologically unrelated drugs in patients with prior history of allergy to other agents. It therefore seems reasonable to avoid cephalosporins in patients with past history of documented anaphylaxis or immediate hypersensitivity reaction to the penicillins, even though anaphylaxis to cephalosporins is very rare. However, cephalosporins may be given to patients with other types of allergy (drug fever, morbilliform rash, etc) to the penicillins (2, 23).

Nephrotoxicity is rare with any of the cephalosporins except cephaloridine, which predictably produces dose-related proximal tubular damage in the kidney (23, 60). The other cephalosporins can produce similar effects in experimental animals, but only at much higher doses than cephaloridine (23). These agents can also rarely produce interstitial nephritis, presumably on an allergic basis (61). There has been a great deal of controversy concerning the ability of cephalosporins to potentiate aminoglycoside nephrotoxicity. Early uncontrolled clinical studies were inconclusive and animal studies showed that cephalosporins protected against, rather than potentiated, aminoglycoside nephrotoxicity (23, 61–63). A recent prospective blind study, however, strongly suggests that nephrotoxicity is more common after cephalosporin-aminoglycoside than after penicillin-aminoglycoside adminstration in man (64).

Hematologic abnormalities are rare in patients receiving the cephalosporins. Although a few patients on cephalosporins develop positive Coombs tests, hemolytic anemia rarely occurs. Reversible leukopenia and thrombocytopenia are also infrequent (55). The cephalosporins interfere with platelet function when present in high concentrations in vitro, but the clinical significance of this observation remains to be proven (65). Patients receiving broad spectrum cephalosporins may develop prolonged prothrombin times, which are reversible with vitamin K administration and are presumably due to the impaired production of vitamin K that results from the suppression of gut flora by these agents (56, R. Kammer, personal communication). Transient elevations in serum transaminase and alkaline phosphatase have been seen in some patients receiving cephalosporins, but this appears to be of no clinical significance (66).

Cephalosporins and Cephamycins Currently Available

PARENTERAL AGENTS

Cephalothin Cephalothin was the first clinically available cephalosporin. It does not produce significant serum or urine levels after oral ingestion, is quite painful when administered intramuscularly, and frequently causes phlebitis after intravenous use. Its activity against β-lactamase-producing *S. aureus* is greater than or equal to that of the other cephalosporins. It is also effective against most other gram-positive organisms, except enterococci. Most strains of *P. mirabilis,* salmonella, shigella, *E. coli,* and *Klebsiella pneumoniae* are susceptible in vitro, whereas indole-positive *Proteus* sp., acinetobacter, pseudomonas, and serratia are resistant (67). Because of rapid renal excretion and resultant short serum half-life, cephalothin should be administered intravenously every four hours and the total daily dose is usually 3–8 g per day (2, 32). Larger doses (up to 16–24 g per day) have been used safely for severe infections.

Cephaloridine The spectrum of antibacterial activity of cephaloridine is similar to that of cephalothin. Its chief differences are its pharmacologic properties and toxicity potential. Unlike other cephalosporins, cephaloridine is not appreciably secreted by the renal tubules; thus the administration of probenicid does not result in enhanced serum concentrations of this drug. Cephaloridine is well tolerated intramuscularly, penetrates relatively well into the CSF, but does not produce high biliary levels (23). The major drawback of cephaloridine is its potential for dose-related nephrotoxicity (60) due to proximal tubular damage. For this reason, the dose of cephaloridine should be limited to 4 g per 24 hours and reduced or avoided altogether if there is preexisting renal insufficiency. With the introduction of other cephalosporins, such as cefazolin, which are also well tolerated intramuscularly, there presently appears to be little indication for using cephaloridine.

Cephacetrile Cephacetrile is a first-generation parenteral cephalosporin unavailable in the US. Doses and indications are the same as for cephalothin (2).

Cephapirin Cephapirin is similar to cephalothin in spectrum of activity, pharmacokinetic properties, and toxic potential (2). Clinical experience with this drug is not as extensive as with cephalothin; no clinically significant differences between the two agents have yet been demonstrated.

Cephradine Cephradine is the only cephalosporin commercially available in both an oral and a parenteral form. The structure, pharmacologic properties, recommended doses, and indications for oral use resemble those for cephalexin (32, 34); neither cephalexin nor cephradine has as broad a spectrum of activity as cefaclor. When used intramuscularly, peak levels may be lower than those achieved after oral administration. Indications and doses for intravenous cephradine are similar to those for cephalothin.

Cefamandole Cefamandole was the first cephalosporin with a truly expanded spectrum of activity to be released in the US. In comparison to cephalothin, it is more active in vitro against *Enterobacter* sp., *Proteus inconstans,* the newly classified *Morganella morganii, Providencia rettgeri, P. stuartii,* some cephalothin-resistant *E. coli* and *K. pneumoniae* (11), and beta-lactamase-producing *H. influenzae.* It has been used successfully in vivo for infections due to these organisms, but the potential for the emergence of resistance in strains of enterobacter remains a concern (68–70). Cefamandole is more active than cefoxitin against gram-positive cocci but much less active than cefoxitin against *Bacteroides fragilis* (30, 31). Cefamandole is usually given every 6 hours intravenously in doses of 2–12 g per day. Intramuscular injections are painful, but the drug can be administered by this route if necessary.

Cefoxitin Cefoxitin is not a cephalosporin but a member of the cephamycin family, a group of structurally similar compounds derived from *Streptomyces* species. Its spectrum of activity is significantly broader than that of cephalothin except against gram-positive cocci. It is active in vitro against many indole-positive *Proteus* sp., *Providencia stuartii,* some strains of *Serratia marcescens,* and many cephalothin-resistant *E. coli* and *K. pneumoniae* (26, 27, 71). It is also active against *Neisseria* sp. including β-lactamase-producing, penicillin-resistant strains of *N. gonorrheae* (72). Cefoxitin is the most active of the currently available cephalosporins against *B. fragilis* and is generally active against other anaerobes as well (30, 31). Of the currently available cephalosporin-cephamycin compounds, cefoxitin appears to be the most resistant to hydrolysis by the β-lactamases produced by gram-negative organisms. Pharmacologic parameters resemble those of cephalothin although peak serum levels of cefoxitin may be slightly higher (2, 26). The suggested dose is 2 g given every 4–6 hours; because IM injections are painful, it is best to give cefoxitin intravenously.

Cefuroxime Cefuroxime is a parenteral cephalosporin presently unavailable in the US. When compared with other available cephalosporins, it displays increased resistance to hydrolysis by β-lactamases found in gram-negative organisms; however, it is less resistant to hydrolysis than is cefoxitin (16, 18, 24). Its in vitro activity is roughtly similar to that of cefamandole, and it is effective against penicillin-resistant gonococci (2, 73).

ORAL AGENTS

Cephaloglycine Cephaloglycine is an early cephalosporin that has lost its place in today's antibiotic armanentarium. It is an oral agent that achieves adequate urinary concentrations but not adequate serum concentrations. Other oral agents are superior either because of a wider spectrum of activity or more favorable pharmacologic properties.

Cephalexin Cephalexin has a spectrum of activity similar to cephalothin. It is well absorbed per os (PO) although it commonly causes mild diarrhea (32). It is excreted unchanged into the urine where very high concentrations are found. Peak serum levels after 500 mg PO are 13–18 μg/ml (23, 32); cephalexin is effective in the therapy of soft tissue infections caused by susceptible organisms. The usual dose is 1–4 g per day in divided doses every 6 hours.

Cefaclor Cefaclor is a new oral agent structurally similar to cephalexin except that a chlorine atom is substituted for the methyl group on the dihydrothiazine ring. The resulting molecule is less stable than cephalexin. In vitro testing shows that cefaclor is somewhat more active than cephalexin, cephaloglycine, and cephradine against gram-positive cocci, but it is more susceptible to hydrolysis by staphylococcal β-lactamase (74, 75). Cefaclor is also slightly more active than the other available oral agents against *E. coli, P. mirabilis, K. pneumoniae, S. typhi,* and *N. gonorrheae,* but not against enterococci, *Enterobacter* sp., or indole-positive *Proteus* sp. Perhaps most important are the observations that cefaclor is efficacious against *H. influenzae,* including β-lactamase-producing strains (76, 77). The pharmacokinetics of cefaclor in patients with normal renal function are similar to cephalexin, but in patients with impaired renal function the half-life of cefaclor is not nearly as prolonged as that of cephalexin and dosage adjustment is usually not required (78, 79).

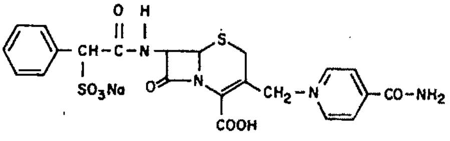
Cefadroxil Cefadroxil is a new semisynthetic cephalosporin that is well absorbed from the gastrointestinal tract even when administered during a meal (80). Peak serum levels may be slightly less than or equal to those achieved with cephalexin but the serum concentration is maintained at a higher level beginning 90 minutes after the dose and continuing through 12 hours (80–82). Although almost all the drug is excreted into the urine, it is cleared more slowly than cephalexin, producing prolonged high urine and serum concentrations. The antibacterial spectrum of cefadroxil is similar to that of cephalexin, although in one study of experimental infections of mice cefadroxil was more protective against *S. aureus* and *S. pneumoniae* (83).

Cephalosporins Under Investigation

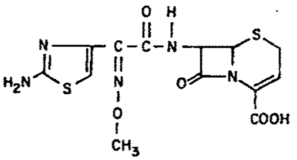
Efforts to create improved β-lactam drugs by manipulations in and around the β-lactam ring have led to the development of β-lactam antibiotics with improved pharmacokinetic properties and/or an expanded spectrum of activity. Several of the more promising of these recent developments are discussed below.

Cefotaxime Cefotaxime (HR 756) is clinically available in Europe, but remains under investigation in the US. It is a semisynthetic cephalosporin whose spectrum of activity against gram-negative bacilli is superior to that of other currently available cephalosporins including cefamandole, cefoxitin, and cefuroxime. It is active in vitro against most *E. coli,* klebsiella, enterobacter, proteus, providencia, citrobacter, and many serratia, even when such organisms are cephalothin or aminoglycoside resistant (84–86). The presence of a methoxy group on the side chain at the 7 position (Figure 2) provides resistance to hydrolysis by many β-lactamases (88) produced by gram-negative bacilli. Cefotaxime is equal to or more active than the older cephalosporins against β-lactamase-producing *H. influenzae* and *N. gonorrheae* (84, 89, 90). Its activity against *P. aeruginosa* is similar to that of carbenicillin and ticarcillin but achievable levels may not be comparable (85, 86, 91–93). It is somewhat less effective than cephalothin or cefamandole against *S. aureus* (84, 87). Its activity against *B. fragilis* is less than that of cefoxitin and it is less active against these organisms than against other anaerobic species (86, 94, 96). Cefotaxime contains an acetyl group at the 3 position (Figure 2); hence it undergoes in vitro desacetylation with resultant loss of activity (87). The clinical significance of this metabolic degradation must be carefully investigated, but it has not yet been a major problem in clinical trials. Early clinical trials suggest that cefotaxime may provide CSF concentrations adequate to treat meningitis (95).

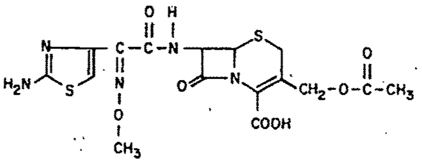
Ceftizoxime Ceftizoxime (FK 749) is the desacetoxymethyl derivative of cefotaxime (Figure 2) (84). Its activity against *H. influenzae, N. gonorrheae,* and enterobacteriaceae is similar to or greater than that of cefotaxime and it is more active against *Serratia marcescens.* It is also more active than



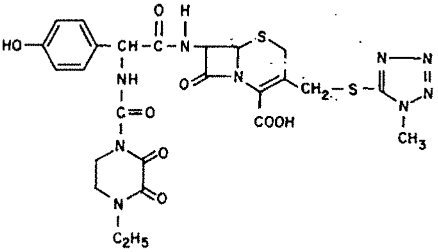
Cefsulodin (SCE-129)



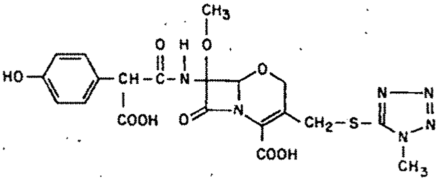
Ceftizoxime (FK 749)



Cefotaxime (HR 756)



Cefoperazone (T 1551)



Moxalactam (LY 127935)

Figure 2 Cefotaxime is a semisynthetic cephalosporin whose spectrum of activity against gram-negative bacilli is superior to that of other currently available cephalosporins including cefamandole, cefoxitin, and cefuroxime.

cefoxitin when tested against *B. fragilis* (93). The level of resistance of ceftizoxime to hydrolysis by penicillinases and cephalosporinases is comparable to that of cefoxitin and cefotaxime; it is slightly more stable than cefuroxime, and much more stable than cefamandole or cefoperazone (15).

Cefoperazone The results of in vitro testing of cefoperazone (T-1551) indicate that, although it is somewhat less active than the above two agents, and about equal to cefoxitin and cefamandole against many gram-negative bacilli, it is more effective against *P. aeruginosa*, inhibiting 50% of strains at a concentration of 5–6.2 μg/ml and 90% of strains at 16–50 μg/ml (85, 96, 97). It appears to be very effective against β-lactamase-positive *H. influenzae* but is less active than cephalothin against *S. aureus* (96).

Cefsulodin Cefsulodin (SCE 129) also offers promise as an antipseudomonal antibiotic, inhibiting 50% and 90% of strains at 3.1 μg/ml and 50 μg/ml respectively (85, 98, 99). It is less active than other cephalosporins against *H. influenzae* (100), enterobacteriaceae, and streptococci (97), close to cefoxitin in effectiveness against many anaerobes (94), and exhibits good antistaphylococcal activity.

Ceforanide Ceforanide is a new semisynthetic, parenteral cephalosporin with a broad spectrum of activity comparable to cefamandole and cefoxitin (101–103). The chief advantage of ceforanide appears to be its pharmacologic properties. Its serum half-life is 2.6–2.8 hours, longer than for other cephalosporins. It is not metabolized, is removed almost entirely by the kidneys, and is bound to protein somewhat less than cefazolin (104, 105). It is well tolerated when given intramuscularly and has been used successfully in clinical trials for cellulitis, pneumonia, and other infections even when given only every 12 hours (106, 107).

Moxalactam Moxalactam (LY 127935) is a 1-oxa-β-lactam that differs in structure from the cephalosporins in that the sulfur atom in the dihydrothiazine ring has been replaced by oxygen (Figure 2). It is the first totally synthetic β-lactam antibiotic and technically, it is not a true cephalosporin. However, like the cephamycins, it shares so many properties with the cephalosporins that it is convenient to consider it in the same context. Moxalactam is highly active in vitro against many enterobacteriaceae including *Enterobacter* sp., indole-positive *Proteus* sp., and *Serratia* sp. Its activity against *P. aeruginosa* approximates that of cefotaxime. For many organisms, it is roughly equal in efficacy to cefotaxime and more active than cefoxitin, cefamandole, or cefuroxime (85–87, 108–111). It is resistant to hydrolysis by almost all gram-negative β-lactamases (112, 113). It is as active against *H. influenzae* and *N. gonorrheae* as cefotaxime, cefoperazone, and ceftizoxime; more active than cefotaxime and cefoxitin against *B. fragilis;* but less active than cephalothin and cefamandole against staphyloccoci and streptococci (86, 94, 111, 114). Moxalactam does not undergo significant in vivo metabolism, and early clinical trials suggest that it provides CSF concentrations adequate to treat gram-negative meningitis.

Conclusion

The cephalosporins (including the closely related cephamycins and their synthetic cousins) represent a group of antimicrobial agents that elucidates clearly the advances in our present understanding of the structural and functional relationships of therapeutic agents. This knowledge has led to the development of many drugs with both pharmacokinetic and therapeutic advantages over the older members of this class. Developing improved drugs by molecular manipulation is a most exciting concept, but it poses problems for the practitioner who is faced with a bewildering array of drugs and whose choices will not be made easier by the rapid proliferation of new cogeners. Indeed, it is only with a thorough understanding of the basic properties of these drugs and via knowledge of their potential therapeutic advantages and disadvantages that the clinician can hope to make a rational choice. For a class of drugs as extensive and as rapidly evolving as the cephalosporins, this represents a major challenge. The benefits to be derived from this process, however, seem worth the effort!

Literature Cited

1. Abraham, E. P. 1962. The cephalosporins. *Pharm. Rev.* 14:473–500
2. Murray, B. E., Moellering, R. C. Jr. 1979. The cephalosporin and cephamycin antibiotics: a status report. *Clin. Ther.* 2:155–80
3. Tomasz, A. 1979. The mechanism of the irreversible antimicrobial effects of penicillins. *Ann. Rev. Microbiol.* 33: 113–37
4. Strominger, J. L. 1977. How penicillin kills bacteria: a short history. In *Microbiology—1977*, ed. D. Schlessinger, pp. 177–81. Washington DC: Am. Soc. Microbiol.
5. Spratt, B. G. 1975. Distinct penicillin-binding proteins involved in the division, elongation, and shape of *Escherichia coli. Proc. Natl. Acad. Sci. USA* 72:2999–3003
6. Tomasz, A., Holtje, J. V. 1977. Murein hydrolases and the lytic and killing action of penicillin. See Ref. 4, pp. 209–91
7. Murray, B. E., Moellering, R. C. 1978. Patterns and mechanisms of antibiotic resistance. *Med. Clin. North Am.* 62:899–923
8. Leive, L. 1974. The barrier function of the gram negative envelope. *Ann. NY Acad. Sci.* 235:109–29
9. Curtis, N. A., Orr, D., Ross, G. W., Boulton, M. G. 1979. Affinities of penicillin and cephalosporins for the penicillin-binding proteins of *Escherichia coli* K12 and their antibacterial activity. *Antimicrob. Agents Chemother.* 16:533–39
10. Boman, H. G., Nordstrom, K., Normark, S. 1974. Penicillin resistance in *E. coli* K12: Synergism between penicillinases and a barrier in the outer part of the envelope. *Ann. NY Acad. Sci.* 235:569–86
11. Moellering, R. C. Jr. 1978. Cefamandole—a new member of the cephalosporin family. *J. Infect. Dis.* 137 (Suppl):52–59
12. Murray, B. E., Moellering, R. C. Jr. 1980. Antimicrobial agents in pulmonary infections. *Med. Clin. North Am.* 64:319–41
13. Richmond, A. S., Simberkoff, M. S., Schaefler, S., Rahal, J. J. 1977. Resistance of *Staphylococcus aureus* to semisynthetic penicillins and cephalosporins. *J. Infect. Dis.* 135:108–12
14. Goldman, P. L., Petersdorf, R. G. 1980. Importance of β-lactamase inactivation in treatment of experimental endocarditis caused by *Staphylococcus aureus. J. Infect. Dis.* 141:331–37

14a. Richmond, M. H., Sykes, R. B. 1973. The β-lactamases of gram-negative bacteria and their possible physiological role. *Adv. Microb. Physiol.* 9:31–88

15. Kojo, H., Nishida, M., Goto, S., Kuwahara, S. 1979. Antibacterial activity of ceftizoxime (FK749), a new cephalosprin, against cephalosporin resistant bacteria, and its stability to β-lactamase. *Antimicrob. Agents Chemother.* 16:539–53
16. Neu, H. C., Fu, K. P. 1978. Cefuroxime, a beta-lactamase-resistant cephalosporin with a broad spectrum of gram positive and negative activity. *Antimicrob. Agents Chemother.* 13:657–64
17. Farrar, W. E., O'Dell, N. M. 1977. Beta lactamase resistance of newer cephalosporins and antimicrobial effectiveness against gram negative bacilli. *Infection* 5:224–27
18. Richmond, M. H., Wotton, S. 1976. Comparative study of seven cephalosporins: susceptibility to beta lactamases and ability to penetrate the surface layers of *Escherichia coli. Antimicrob. Agents Chemother.* 10:219–22
19. Aswapokee, N., Aswapokee, P., Fu, K. P., Neu, H. C. 1978. In vitro activity and beta lactamase stability of BL-S786 compared with those of other cephalosporins. *Antimicrob. Agents Chemother.* 14:1–5

20. Fu, K. P., Neu, H. C. 1978. Beta-lactamase stability of HR756, a novel cephalosporin compared to that of cefuroxime and cefoxitin. *Antimicrob. Agents Chemother.* 14:322–26
21. Onishi, H. R., Daoust, D. R., Zimmerman, S. B., Hendlin, D., Stapley, E. O. 1974. Cefoxitin, a semisynthetic cephamycin antibiotic: resistance to β-lactamase inactivation. *Antimicrob. Agents Chemother.* 5:38–48
22. John, J. F., McNeil, W. F. 1980. Activity of cephalosporins against methicillin susceptible and methicillin resistant, coagulase negative staphylococci: minimal effect of β-lactamase. *Antimicrob. Agents Chemother.* 17:179–83
23. Moellering, R. C., Swartz, M. N. 1976. The newer cephalosporins. *N. Engl. J. Med.* 294:24–28
24. Neu, H. 1978. Another cephalosporin! *Ann. Intern. Med.* 89:719–20
25. Jorgensen, J., Alexander, G. A. 1978. Comparative activities of selected betalactam antibiotics against *Haemophilus influenzae. Antimicrob. Agents Chemother.* 13:342–43
26. Brumfitt, W., Kosmidis, J., Hamilton-Miller, J. M. T., Gilchrist, J. N. G. 1974. Cefoxitin and cephalothin: antimicrobial activity, human pharmacokinetics, and toxicity. *Antimicrob. Agents Chemother.* 6:290–99
27. Wallick, H., Hendlia, D. 1974. Cefoxitin, a semisynthetic cephamycin antibiotic: susceptibility studies. *Antimicrob. Agents Chemother.* 5:25–32
28. Jones, R. N., Fuchs, P. C., Gavan, T. L., Gerlach, E. H., Barry, A. L., Thornsberry, C. 1977. Cefuroxime, a new parenteral cephalosporin: collaborative *in vitro* susceptibility comparison with cephalothin against 5887 clinical bacterial isolates. *Antimicrob. Agents Chemother.* 12:47–50
29. Verbist, L. 1976. Comparison of the antibacterial activity of 9 cephalosporins against *Enterobacteriaciae* and nonfermentative gram negative bacilli. *Antimicrob. Agents Chemother.* 10:657–63
30. Tally, F. P., Jacobus, N. V., Bartlett, J. G., Gorbach, S. L. 1975. Susceptibility of anaerobes to cefoxitin and other cephalosporins. *Antimicrob. Agents Chemother.* 7:128–32
31. Sutter, V. L., Finegold, S. M. 1975. Susceptibility of anaerobic bacteria to carbenicillin, cefoxitin and related drugs. *J. Infect. Dis.* 131:417–22
32. Nightingale, C. H., Greene, D. S., Quintiliani, R. 1975. Pharmcokinetics and clinical use of cephalosporin antibiotics. *J. Pharm. Sci.* 64:1899–1927
33. Neu, H. C. 1978. Comparison of the pharmacokinetics of cefamandole and other cephalosporin compounds. *J. Infect. Dis.* 137 (Suppl.):S80–S87
34. Brogard, J. M., Comte, F., Pinget, M. 1978. Pharmacokinetics of cephalosporin antibiotics. *Antibiot. Chemother.* 25:123–62
35. Andriole, V. T. 1978. Pharmacokinetics of cephalosporins in patients with renal failure or reduced renal function. *J. Infect. Dis.* 137(Suppl.):S88–S97
36. Guerrero, I. C., MacGregor, R. R. 1979. Comparative penetration of various cephalosporins into inflammatory exudate. *Antimicrob. Agents Chemother.* 15:712–15
37. Barza, M., Weinstein, L. 1975. Penetration of antibiotics into fibrin loci in vivo. I. Comparison of penetration of ampicillin into fibrin clots, abscesses and interstitial fluid. *J. Infect. Dis.* 129:59–64
38. Carbon, C., Contrepois, A., Brion, N., Lamotte-Barillon, S. 1977. Penetration of cefazolin, cephaloridine, and cefamandole into interstitial fluid in rabbits. *Antimicrob. Agents Chemother.* 11:594–98
39. Gerding, D. N., Peterson, L. R., Legler, D. C., Hall, W. H., Schierl, E. A. 1978. Ascitic fluid cephalosporin concentrations: influence of protein binding and serum pharmacokinetics. *Antimicrob. Agents Chemother.* 14:234–39
40. Gerding, D. N., Kromhout, J. P., Sullivan, J. J., Hall, W. H. 1976. Antibiotic penetrace of ascitic fluid in dogs. *Antimicrob. Agents Chemother.* 10:850–55
41. Nightingale, C. H., Klimek, J. J., Quintiliani, R. 1980. Effect of protein binding on the penetration of nonmetabolized cephalosporins into atrial appendage and pericardial fluids in open heart surgical patients. *Antimicrob. Agents Chemother.* 17:595–98
42. Lerner, P. I. 1971. Penetration of cephaloridine into cerebrospinal fluid. *Am. J. Med. Sci.* 162:321–26
43. Korzeniowski, O. M., Carvalho, E. M., Rocha, H., Sande, M. A. 1978. Evaluation of cefamandole therapy of patients with bacterial meningitis. *J. Infect. Dis.* 137(Suppl.):S169–79
44. Steinberg, E. A., Overturf, G. D., Wilkins, J., Baraff, L. J., Streng, J. M., Leedom, J. M. 1978. Failure of cefamandole in treatment of meningitis due to *Haemophilus influenzae* type b. *J. Infect. Dis.* 137(Suppl.):S180–86

45. Hubert, G., Leroy, A., Roger, J. P., Cherubin, C. 1980. Cefoxitin concentrations in the cerebrospinal fluids of patients with meningitis. *Antimicrob. Agents Chemother.* 17:675–78
46. Massip, P., Kitzis, M. D., Tran, V. T., Armengaud, M. J., Armengaud, M. 1979. Penetration of cefoxitin into cerebrospinal fluid of dogs with and without experimental meningitis. *Rev. Infect. Dis.* 1:132–33
47. Nair, S. R., Cherubin, C. E., Weinstein, M. 1979. Penetration of cefoxitin into cerebrospinal fluid and treatment of meningitis caused by gram negative organisms. *Rev. Infect. Dis.* 1:134–42
48. Muller, C., Netland, A., Dawson, A. F., Andrew, E. 1980. The Penetration of cefuroxime into the cerebrospinal fluid through inflamed and non-inflamed meninges. *J. Antimicrob. Chemother.* 6:279–283
49. Friedrich, H., Haensel-Friedrich, G., Langmaak, H., Daschner, F. D. 1980. Investigations of cefuroxime levels in the cerebrospinal fluid of patients with and without meningitis. *Chemotherapy* 26:91–97
50. Ram, M. D., Watanatittan, S. 1976. Biliary excretion and concentration of cefazolin. *Am. J. Gastroenterol.* 66: 540–45
51. McLeish, A. R., Strachan, C. J., Powis, S. J., Wise, R., Bevan, P. G. 1977. The influence of biliary disease on the excretion of cefazolin in human bile. *Surgery* 81:426–30
52. Waterman, N. G., Eickenberg, H. U., Scharfenberger, L. 1976. Concentration of cefamandole in serum, interstitial fluid, bile, and urine. *Antimicrob. Agents Chemother.* 10:733–35
53. Ratzan, K. R., Baker, H. B., Lauredo, I. 1978. Excretion of cefamandole, cefazolin, and cephalothin into T-tube bile. *Antimicrob. Agents Chemother.* 13:985–87
54. Berger, S., Ernst, E. C., Barza, M. 1976. Comparative incidence of phlebitis due to buffered cephalothin, cephapirin, and cefamandole. *Antimicrob. Agents Chemother.* 9:575–79
55. Petz, L. D. 1978. Immunologic cross-reactivity between penicillins and cephalosporins: a review. *J. Infect. Dis.* 137(Suppl.):S74–S87
56. Calderwood, S. B., Moellering, R. C. Jr. 1980. Common adverse effects of antibacterial agents on major organ systems. *Surg. Clin. North Am.* 60:65–81
57. Levine, B. B. 1973. Antigenicity and cross reactivity of the penicillins and cephalosporins. *J. Infect. Dis.* 128 (Suppl.):S364–66
58. Kuwahara, S., Mine, Y., Nishida, M. 1971. Immunogenicity of cefazolin. *Antimicrob. Agents Chemother. 1970,* pp. 374–79
59. Petz, L. D. 1971. Immunologic reactions of humans to cephalosporins. *Postgrad. Med. J.* (Feb. Suppl.), pp. 64–69
60. Appel, G. B., Neu, H. C. 1977. The nephrotoxicity of antimicrobial agents. *N. Engl. J. Med.* 296:663–70, 722–28, 784–86
61. Barza, M. 1978. The nephrotoxicity of cephalosporins: an overview. *J. Infect. Dis.* 137(Suppl.):S60–S73
62. Dellinger, P., Murphy, T., Pinn, V., Barza, M., Weinstein, L. 1976. Protective effect of cephalothin against gentamicin-induced nephrotoxicity in rats. *Antimicrob. Agents Chemother.* 9: 172–78
63. Wold, J. S., Turnipseed, S. A., Broddle, W. D., Owne, N. V. 1977. Effect of cefamandole nafate on the toxicity of tobramycin. *Antimicrob. Agents Chemother.* 12:465–69
64. Wade, J. C., Smith, C. R., Petty, B. G., Lipsky, J. J., Conrad, G., Ellner, J., Lietman, P. S. 1978. Cephalothin plus an aminoglycoside is more nephrotoxic than methicillin plus an aminoglycoside. *Lancet* 2(8090):604–6
65. Natelson, E. A., Brown, C. H. III, Bradshaw, M. W., Alfrey, C. P. Jr., Williams, T. W. Jr. 1976. Influence of cephalosporin antibiotics on coagulation and platelet function. *Antimicrob. Agents Chemother.* 9:91–93
66. Barza, M., Miao, P. V. W. 1977. Antimicrobial spectrum, pharmacology, and therapeutic use of antibiotics. Part 3: cephalosporins. *Am. J. Hosp. Pharm.* 34:621–29
67. Washington, J. A. 1976. The in vitro spectrum of the cephalosporins. *Mayo Clin. Proc.* 51:237–50
68. Rodriguez, W. J., Ross, S., Khan, W. N., Goldenberg, R. 1978. Clinical and laboratory evaluation of cefamandole in infants and children. *J. Infect. Dis.* 137(Suppl.)S150–54
69. Levine, L. R., McCain, E. 1978. Cefamandole in the treatment of infections due to *Enterobacter and indole-positive Proteus. J. Infect. Dis.* 137(Suppl.): S125–32
70. Findell, C. M., Shervis, J. C. 1976. Susceptibility of *Enterobacter* to cefamandole: evidence for a high mutation rate to resistance. *Antimicrob. Agents Chemother.* 9:970–74

71. Moellering, R. C. Jr., Dray, M., Kunz, L. J. 1974. Susceptibility of clinical isolates of bacteria to cefoxitin and cephalothin. *Antimicrob. Agents Chemother.* 6:320–23
72. Siegel, M. S., Thompson, S. E., Perine, P. L., Brown, S. T., Reynolds, G., Thornsberry, T. 1979. Treatment of uncomplicated gonococcal urethritis with cetoxitin: comparison with penicillin. *Rev. Infect. Dis.* 1:183–87
73. Percival, A., Corkill, J. E., Arya, O. P., Rowlands, J., Alerant, C. D., Rees, E., Annels, E. H. 1976. Penicillinase-producing gonococci in Liverpool. *Lancet* 2:1379–89
74. Shadomy, S., Wagner, G., Carver M. 1977. In vitro activities of 5 oral cephalosporins against aerobic pathogenic bacteria. *Antimicrob. Agents Chemother.* 12:609–13
75. Silver, M. S., Counts, G. W., Zeleznik, D., Turck, M. 1977. Comparison of in vitro antibacterial activity of 3 oral cephalosporins: cefaclor, cephalexin and cephradine. *Antimicrob. Agents Chemother.* 12:591–96
76. Sinai, R., Hammberger, S., Marks, M. I., Pai, C. H. 1978. In vitro susceptibilities of *Haemophilus influenzae* to sulfamethoxazole-trimethoprim and cefaclor, cephalexin, and cephradine. *Antimicrob. Agents Chemother.* 13:861–64
77. Bill, N. J., Washington, J. A. 1977. Comparison of in vitro activity of cephalexin, cephradine, and cefaclor. *Antimicrob. Agents Chemother.* 11:470–74
78. Hodges, G. R., Liu, C., Hinthorn, D. R., Harms, J. L., Dworzack, D. L. 1978. Pharmacologic evaluation of cefaclor in volunteers. *Antimicrob. Agents Chemother.* 13:454–56
79. Santoro, J., Agarwal, B. N., Martinelli, R., Wengir, N., Levison, M. E. 1978. Pharmacology of cefaclor in normal volunteers and in patients with renal volunteers. *Antimicrob. Agents Chemother.* 13:951–54
80. Lode, H., Stahlamnn, R., Koeppe, P. 1979. Comparative pharmacokinetics of cephalexin, cefaclor, cefadroxil, and CGP 9000. *Antimicrob. Agents Chemother.* 16:1–6
81. Humbert, G., Leroy, A., Fillastre, J. P., Godin, M. 1979. Pharmacokinetics of cefadroxil in normal subjects and in patients with renal insufficiency. *Chemotherapy* 25:189–95
82. Prenna, M., Ripa, S. 1980. Serum levels and urinary excretion in humans of BL-S578 (cefadroxil), a new semisynthetic cephalosporin. *Chemotherapy* 26:98–102
83. Ripa, S., Prenna, M. 1979. Laboratory studies with BL-S578 (defadroxil), a new broad spectrum orally active cephalosporin. *Chemotherapy* 25:9–13
84. Greenwood, D., Pearson, N., Ely, A., O'Grady, F. 1980. Comparative in vitro activities of cefotaxime and ceftizoxime (FK 749): new cephalosporins with exceptional activity. *Antimicrob. Agents Chemother.* 17:397–401
85. Hall, W. H., Opfer, B. J., Gerding, D. N. 1980. Comparative activities of the oxa-β-lactam LY127935, cefotaxime, cefoperazone, cefamandole, and ticarcillin against multiple resistant gram negative bacilli. *Antimicrob. Agents Chemother.* 17:273–79
86. Neu, H. C., Aswapokee, N., Fu, K. P., Aswapokee, P. 1979. Antibacterial activity of a new 1-oxa-cephalosporin compared with that of other β-lactam compounds. *Antimicrob. Agents Chemother.* 16:141–49
87. Wise, R., Wills, P. J., Andrews, J. M., Bedford, K. A. 1980. Activity of the cefotaxime (HR756) desacetyl metabolite compared with those of cefotoxine and other cephalosperms. *Antimicrob. Agents Chemother.* 17:84–86
88. Fu, K. P., Neu, H. C. 1979. The comparative β-lactamase resistance and inhibitory activity of 1-oxa-cephalosporin, cefoxitin, and cefotaxime. *J. Antibiot.* 32:909–14
89. Baker, C. N., Thornsberry, C., Jones, R. N. 1980. In vitro antimicrobial activity of cefoperazone, cefotaxime, moxalactam (LY127935), azlocillin, mezlocillin, and other β-lactam antibiotics against *N. gonorrhoeae* and *H. influenzae,* including β-lactamase-producing strains. *Antimicrob. Agents Chemother.* 17:757–61
90. Jorgensen, J. H., Crawford, S. A., Alexander, G. A. 1980. *In vitro* activities of cefotaxime and moxalactam (LY 127935) against *Haemophilus influenzae. Antimicrob. Agents Chemother.* 17:516–17
91. Jacoby, G. A., Sutton, L. 1979. Activity of β-lactam antibiotics against *Pseudomonas aeruginosa* carrying R plasmids determining different β-lactamases. *Antimicrob. Agents Chemother.* 16:243–45
92. Kamimura, T., Matsumoto, Y., Okada, N., Mine, Y., Nishida, M., Goto, S., Kuwahara, S. 1979. Ceftizoxime (FK 749) a new parenteral cephalosporin: in vitro and in vivo antibacterial activities.

Antimicrob Agents Chemother. 16:540–48

93. Fu, K. P., Neu, H. C. 1980. Antibacterial activity of ceftizoxime, a β-lactamase stable cephalosporin. *Antimicrob. Agents Chemother.* 17:583–90
94. Borobio, M. V., Aznar, J., Jimenez, R., Garcia, F., Perea, E. J. 1980. Comparative in vitro activity of 1-oxa-β-lactam (LY127935) and cefoperazone with other β-lactam antibiotics against anaerobic bacteria. *Antimicrob. Agents Chemother.* 17:129–31
95. Belohradsky, B. H., Bruch, K., Geiss, D., Kafetzis, D., Marget, W., Peters, G. 1980. Intravenous cefotaxime in children with bacterial meninitis. *Lancet* 1:61–63
96. Neu, H. C., Fu, K. P., Awsapokee, N., Awsapokee, P., Kung, K. 1979. Comparative activity and β-lactamase stability of cefoperazone, a piperazine cephalosporin. *Antimicrob. Agents Chemother.* 16:150–57
97. Hinkle, A. M., LeBlanc, B. M., Bodey, G. P. 1980. In vitro evaluation of cefoperazone. *Antimicrob. Agents Chemother.* 17:423–27
98. Tsuchiya, K., Kondo, M. 1978. Comparative in vitro activities of SCE 129, Sulbenicillin, Gentamicin, and Dibekacin against *Pseudomonas. Antimicrob. Agents Chemother.* 13:536–39
99. Neu, H. C., Fu, K. P. 1979. In vitro antibacterial activity and β-lactamase stability of SCE 129, a new cephalosporin. *Antimicrob. Agents Chemother.* 15:646–50
100. Bulger, R. R., Washington, J. A. 1980. Effect of inoculum size and β-lactamase production on in vitro activity of new cephalosporins against *Haemophilus* species. *Antimicrob. Agents Chemother.* 17:393–96
101. Weaver, S. S., LeBlanc, B. M., Bodey, G. P. 1979. BL-S786 (Ceforanide), a new parenteral cephalosporin: in vitro studies. *Antimicrob. Agents Chemother.* 15:318–22
102. Aswapokee, N., Aswapokee, P., Fu, K. P., Neu, H. 1978. In vitro activity and β-lactamase stability of BL-S786 compared with those of other cephalosporins. *Antimicrob. Agents Chemother.* 14:1–5
103. Jones, R. N., Thornsberry, C., Barry, A. L., Fuchs, P. C., Gavan, T. L., Gerlach, E. H. 1977. BL-S786, a new parenteral cephalosporin. II. In vitro antimicrobial activity comparison with six related cephalosporins. *J. Antibiot.* 30:584–92
104. Lee, F. H., Pfeffer, M., Van Harken, D. R., Smith, R. D., Hottendorf, G. H. 1980. Comparative pharmacokinetics of ceforanide (BL-S786R) and cefazolin in laboratory animals and humans. *Antimicrob. Agents Chemother.* 17:188–92
105. Smyth, R. D., Pfeffer, M., Glick, A., Von Harken, D. R., Hottendorf, G. H. 1978. Clinical pharmacokinetics and safety of high doses of ceforanide (BL-S786R) and cefazolin. *Antimicrob. Agents Chemother.* 17:615–21
106. Burch, K. H., Pohlod, D., Saravolatz, L. D., Madhavan, T., Kiam, T., Quinn, E. L., Del Busto, R., Cardenas, J., Fisher, E. J. 1979. Ceforanide: in vitro and clinical evaluation. *Antimicrob. Agents Chemother.* 16:386–91
107. Musher, D. M., Fainstein, V., Young, E. J. 1980. Treatment of cellulitis with ceforanide. *Antimicrob. Agents Chemother.* 17:254–57
108. Reimer, L. G., Mirrett, S., Reller, L. B. 1980. Comparison of in vitro activity of moxalactam (LY127935) with cefazolin, amikacin, tobramycin, carbenicillin, piperacillin, and ticarcillin against 420 blood culture isolates. *Antimicrob. Agents Chemother.* 17:412–16
109. Delgado, D. G., Brau, C. J., Cobbs, C. G., Dismukes, W. E. 1980. In vitro activity of LY127935, a new 1-oxa-cephalosporin, against aerobic gram negative bacilli. *Antimicrob. Agents Chemother.* 17:864–68
110. Jones, R. N., Fuchs, P. C., Barry, A. L., Gavan, T. L., Sommers, H. M., Gerlach, E. H. 1980. Cefoperazone (T1551), a new semisynthetic cephalosporin. Comparison with cephalothin and gentamicin. *Antimicrob. Agents Chemother.* 17:743–51
111. Wise, R., Andrews, J. M., Bedford, K. A. 1979. LY127935 a novel oxa-β-lactam: an in vitro comparison with other β-lactam antiobiotics. *Antimicrob. Agents Chemother.* 16:341–45
112. Weaver, S. S., LeBlanc, B. M., Bodey, G. P. 1980. In vitro studies of 1-oxacephalosporin (LY127935), a new betalactam antibiotic. *Antimicrob. Agents Chemother.* 17:92–95
113. Yoshida, T., Matsuura, S., Mayama, M., Kameda, Y., Kuwahara, S. 1980. Moxalactam (6059-S), a novel 1-oxa-β-lactam with an expanded antibacterial spectrum: laboratory evaluation. *Antimicrob. Agents Chemother.* 17:302–12
114. Jones, R. N., Fuchs, P. C., Sommers, H. M., Gavan, T. L., Barry, A. L., Gerland, E. H. 1980. Moxalactam (LY127935), a new semisynthetic 1-

oxa-β-lactam antibiotic with remarkable antimicrobial activity: in vitro comparison with cefamandole and tobramycin. *Antimicrob. Agents Chemother.* 17:750–56

115. O'Callaghan, C. W., Harding, S. M. 1977. The pharmacokinetics of cefuroxime in man in relation to its antibacterial activity. *Proc. R. Soc. Med.* 70(Suppl. 9):4–10
116. Spyker, D. A., Thomas, B. L., Sande, M. A., Bolton, W. K. 1978. Pharmacokinetics of cefaclor and cephalexin: dosage nomograms for impaired renal function. *Antimicrob. Agents Chemother.* 14:172–77
117. Bach, V. T., Khurana, M. M., Thadepalli, H. 1978. In vitro activity of cefaclor against aerobic and anaerobic bacteria. *Antimicrob. Agents Chemother.* 13:210–13
118. Griffith, R. S., Black, H. R., Brier, G. L., Wolny, J. D. 1976. Cefamandole: in vitro and clinical pharmacokinetics. *Antimicrob. Agents Chemother.* 10:814–23
119. Fong, I. W., Ralph, E. D., Engelking, E. R., Kirby, W. M. M. 1976. Clinical pharmacology of cefamandole as compared with cephalothin. *Antimicrob. Agents Chemother.* 9:56–59
120. Foord, R. D. 1976. Cefuroxime: human pharmacokinetics. *Antimicrob. Agents Chemother.* 9:741–47

Ann. Rev. Med. 1981. 32:583–99
Copyright © 1981 by Annual Reviews Inc. All rights reserved

DIFFERENTIAL DIAGNOSIS AND TREATMENT OF PANIC ATTACKS AND PHOBIC STATES

❖7420

Michael R. Liebowitz, M.D. and Donald F. Klein, M.D.

Department of Psychiatry, College of Physicians and Surgeons, Columbia University, and the New York State Psychiatric Institute, 722 West 168th Street, New York, New York 10032

Introduction

Twenty years ago a physician's ability to help patients with severe anxiety (panic) attacks or various types of phobia was quite limited. During the past two decades, however, our understanding of and ability to treat these conditions has increased immensely. This chapter, which begins with an illustrative case, documents this progress in the areas of differential diagnosis, pathophysiology, pharmacotherapy, and behavioral treatment approaches.

Case Report

INTRODUCTION AND CHIEF COMPLAINT The patient was a 32-year-old married woman referred by her internist for psychiatric evaluation and treatment of recurrent "severe anxiety attacks."

PRESENT ILLNESS The patient reported that for the past six months she had been experiencing episodes of severe anxiety. These episodes would start without identifiable precipitants, build up "like a wave" to a level of panic over 15–20 minutes, and gradually subside by the end of an hour. When having these attacks, the patient experienced the following: "My chest tightens, my ears clog, my palms sweat, I have a sense of dread, and it feels like I'm losing control." These episodes occurred in a variety of situations where fast exit was difficult, such as supermarket lines, buses, crowded stores, restaurants, and elevators. In each case the patient would immediately flee the situation, even if it meant, for example, abandoning a

0066-4219/81/0401-0583$01.00

basket full of groceries she was waiting to pay for. Furthermore, she had begun to limit her activities because she feared having a panic attack: she avoided restaurants and department stores, took taxis instead of buses, and shopped at times when there were no lines in the supermarkets.

She had seen her internist several times for this problem. He had prescribed diazepam (Valium®) 10 mg t.i.d. for her anxiety symptoms, which calmed her between episodes but did not block the panic attacks. She was on no other medication. She also denied use of other psychoactive agents or abuse of alcohol.

The patient had been upset and worried by her elderly father's declining health over the past year. She had also suffered a spontaneous abortion eight months before her anxiety symptoms began after developing preeclampsia in her sixth month of pregnancy. Finally, she had experienced several weeks of "unrelenting criticism" by her superior at work immediately prior to the onset of her panic attacks.

Her first panic attack occurred during a business luncheon, from which she had to hastily excuse herself. Because of her boss's constant criticism and refusal to grant her vacation time to visit her ailing father, she quit her job, stayed at home without working for three months, and then took a less demanding job, all without resolution of her symptoms.

She also began to feel "depressed" after several months, with a loss of enthusiasm for life, accompanied by initial insomnia and a 15-lb weight loss. This lasted for two months; however, she had regained the lost weight by the time of her initial psychiatric consultation, although her initial insomnia persisted. In addition, her abilities to enjoy friends, her husband, sex, reading, television, etc were at no time impaired. Furthermore, she denied middle or terminal insomnia, loss of energy, symptoms of agitation or psychomotor retardation, self-reproach or inappropriate guilt feelings, diurnal variation, inability to concentrate or make decisions, or recurrent thoughts about death or suicide.

PAST PSYCHIATRIC HISTORY The patient had experienced one prior attack several years earlier when trapped in a malfunctioning elevator. Otherwise, she had no prior history of psychiatric disorder and had never previously consulted a psychiatrist. She had been despondent for several weeks after her miscarriage, but without vegetative signs and with prompt return to normal function.

SOCIAL HISTORY The patient was the oldest of four siblings. The only family history of psychiatric disorder was a long stay in a sanitarium by her paternal grandmother for "nervous breakdown." The patient described her childhood as "happy." She had no difficulty leaving home for school or summer camp. The family was described as economically secure and socially well placed, and her parents got along reasonably well.

The patient's father is reported to have at time abused alcohol and become verbally abusive to his children, criticizing their intelligence and challenging them to argue back and defend themselves. The patient would "fight back" more than would her siblings, but felt at such times that she was in a "no-win" situation in that her father liked to argue and to outwit his children and would never admit he was wrong. Thus the arguments would continue until the children submitted, which was the only way to regain their father's affection. In general, the patient felt that she and her siblings had to humor her father to maintain his affection and approval, and, if they didn't humor him, they had to stay out of his way.

There appeared to be no unusual problems in adolescence, either academic or social. She attended a highly rated college. She then went on to obtain a graduate degree, married a successful businessman, and finally settled in a large city. She worked for several years at various jobs, finally quitting because she felt male co-workers were being advanced more quickly than she. She then entered the advertising field, during which time she became pregnant for the first time, developed toxemia, and miscarried. She went back to the same job but after several months started feeling unfairly and unrelentingly criticized by her boss. It was at this point that her panic symptoms began.

MEDICAL HISTORY The patient underwent a post-traumatic splenectomy in her mid-20s. The only other pertinent medical history is described in the present illness.

The patient had had a full work-up by her internist, including physical examination and laboratory evaluation immediately prior to her referral; the results were within normal limits.

MENTAL STATUS The patient presented as a slender, fashionably dressed woman looking her stated age. She appeared worried about her symptoms, distressed by their restrictive effect on her life, and bewildered as to why they were happening. She was sad without psychomotor retardation and was without psychotic ideation or cognitive impairment. She related in a cooperative and appropriate manner.

Differential Diagnosis

This case raises many interesting questions pertaining to the differential diagnosis, treatment, and clinically relevant research of anxiety and phobic disorders. In terms of diagnosis, using the Diagnostic and Statistic Manual (DSM-II), this patient would have been diagnosed as having features of an anxiety neurosis, a phobic neurosis, and a depressive neurosis. In DSM-II an anxiety neurosis was characterized by "anxious over-concern"—extending to panic and frequently associated with somatic symptoms such as faintness or palpitations—that would occur under any circumstances. A patient with a phobic neurosis might show similar symptoms, but only when exposed to specific situations or objects, though these were consciously recognized as presenting no real danger. Phobias were generally attributed to fears displaced to the phobic situation or object from some other object of which the person was unaware.

Clinical research findings of the past two decades, however, now permit us to more precisely dissect these anxiety and phobic conditions. For example, the patient discussed above began to have brief episodes of intense anxiety or panic six months before presenting for psychiatric treatment. While first occurring at a business luncheon, the panic attacks soon began to happen in supermarket lines, buses, crowded stores, restaurants, and elevators. Whenever they occurred, she was forced to precipitously flee the situation before her anxiety symptoms would abate.

Within a short time the patient began to feel anxious before entering the above-mentioned situations; that is, she began to worry that she might become panicky. She was thus experiencing anticipatory anxiety when she had to go to a supermarket, restaurant, bus, or elevator. The third and final step was to begin to avoid the feared situations, to develop phobic avoidance of places in which panic attacks had occurred. These seem to include any situation in which immediate exit was difficult or access to trusted persons restricted.

The crucial distinction to be made is that the core event was the spontaneous panic attacks, which led to the development of anticipatory anxiety and then phobic avoidance. That the panic attack is qualitatively distinct from, and not just an intensification of, the anticipatory anxiety is supported by their differential pharmacological responses. Klein (1) demonstrated that the panic attacks themselves could be blocked by the tricyclic antidepressant imipramine, which, however, did not affect the anticipatory anxiety. Benzodiazepines and other minor tranquilizers, on the other hand, ameliorated the anticipatory anxiety but had no effect on the panic attacks themselves. Thus, panic attacks appear to represent a form of anxiety biologically distinct from anticipatory (or "free-floating") anxiety.

This in turn leads to a more logical classification of phobic disorders, with a separation of those involving spontaneous panic attacks from those that do not. In DSM-III, the former will be classified as Panic Disorder or Agoraphobia with Panic Attacks, depending on whether an individual with spontaneous panic attacks also has a widespread inability to travel outside the home. "Agoraphobia," a fear of open or public places, is most often actually a fear of leaving home, based in turn on a fear of having a panic attack. The panic attacks themselves are often episodic, and may clear spontaneously after 6–12 months. However, they tend to recur and occasionally they become chronic. In addition, the agoraphobic behavior often becomes chronic because after a time the patient may give up trying to enter feared situations.

Other types of phobias that do not involve spontaneous panic attacks are listed in DSM-III under the headings of Agoraphobia without Panic Attacks, Social Phobia, and Simple Phobia. Agoraphobia, as mentioned previously, involves a fear of open or public places away from home. Although often preceded by panic attacks, it occasionally occurs in their absence. Social Phobia is defined in DSM-III as a persistent, irrational fear of, and compelling desire to avoid, a situation in which the individual is exposed to possible scrutiny by others and fears that he or she may act in a way that will be humiliating or embarrassing. Examples of Social Phobia are fears of public speaking or performing in public, using public lavatories, eating in public, and writing in the presence of others. Simple Phobia involves an irrational fear of, and desire to avoid, objects or situations not included in the definitions of Agoraphobia and Social Phobia. These often involve animals, particularly dogs, snakes, insects, and mice, or situations such as closed spaces (claustrophobia) or heights (acrophobia). As will become clear later in the discussion, the presence or absence of spontaneous panic attacks has crucial treatment implications.

Physical disorders such as withdrawal from sedatives, hypoglycemia, pheochromocytoma, and hyperthyroidism may cause symptoms similar to a panic attack and must be ruled out when making the diagnoses of Panic Disorder and Agoraphobia with Panic Attacks. These diagnoses are also not made in DSM-III if the panic attacks are symptomatic of another mental disorder, such as Schizophrenia, Major Depressive Disorder, or Somatization Disorder.

A number of features complicate the clinical picture and serve to obscure the diagnosis of Panic Disorder. As first noted by Roth (2), some patients with phobic anxiety also manifest symptoms of depersonalization or derealization, which may occur at times other than during a panic attack. There may also be a relationship between separation anxiety in childhood and panic attacks, since between 20% (outpatients) and 50% (inpatients) of agoraphobic adults who present for treatment have a history of separation anxiety in childhood, often manifested as difficulty attending school (1).

Adult patients with Agoraphobia with Panic Attacks, who may present as housebound, desperately clinging, and fearful of being alone, have often been regarded as infantile personalities with severe, unmet dependency needs. The tendency has been to regard their symptoms as evidence of a failure to master the early stages of psychological development involving object constancy and separation-individuation. Long courses of psychoanalytically based treatment were often prescribed, which, in the light of recent findings, may not be the treatment of choice.

Other complications of panic attacks are alcohol or sedative abuse. These may involve attempts at self-medication, which diminish anticipatory anxiety but have no effect on blocking panic attacks. Hence the dosages are constantly increased until physical or psychological dependence supervenes. An inquiry about panic attacks should always be included in the evaluation of alcohol- or sedative-abusing patients.

Panic attacks may also occur in conjunction with a depressive illness (3). At times the panic attacks may be the original symptom, with depression supervening. Most often this involves demoralization in response to the travel restrictions and attendant loss of self-esteem, and lacks many of the features of a full-blown depressive syndrome On the other hand, the illness may start with a depression and be complicated by the subsequent development of panic attacks. According to DSM-III guidelines, if panic attacks occur solely in the context of a depressive episode, then only a diagnosis of Depressive Disorder is made. However, if the panic attacks predate the depression, as was the case with the patient discussed above, then a diagnosis of Panic Disorder or Agoraphobia with Panic Attacks is warranted in addition to that of a Depressive Disorder. However, while demoralized about her functional impairment, this patient lacked sufficient symptoms to meet DSM-III criteria for a Major Depressive Disorder. She was therefore given the secondary diagnosis of Adjustment Disorder with Depressed Mood.

Several recent reports have also linked anxiety-panic syndromes to cardiovascular disorders such as hyperdynamic beta-adrenergic state (4, 5) and mitral valve prolapse syndrome (MVPS) (3). While hyperdynamic beta-adrenergic circulatory state was originally described in patients who had attacks of palpitations, chest discomfort, or tachycardia after exercise or stress (4), it has more recently been reported as brief episodes of nervousness or anxiety, often occurring without physical or emotional precipitants, that can be reproduced by isoproterenol infusions and blocked by maintenance propranolol (5). These findings suggest that some panic-attack patients may have increased beta-adrenergic receptor site reactivity.

Patients with MVPS are prone to extrasystoles, cardiac awareness, tachycardia, palpitation, syncope, fatigue, dyspnea, and atypical chest pain (3). The principal anatomic defect is a redundance of the myxomatous connective tissue of the mitral valve which causes the valve to billow or prolapse into the left atrium during systole. The diagnosis is made on the basis of characteristic auscultatory findings and by echocardiogram. In one series of 17 patients with panic attacks who received full cardiac evaluations, six had definite and two had probable evidence of MVPS (3), which suggests a strong relationship between the two disorders. However, the incidence of MVPS in the general population is uncertain and at first we were not sure that it differed significantly from that found in Pariser et al's sample. For example, one recent study of 100 presumably healthy young females reported that 17% showed phonocardiographic and 21% echocardiographic evidence of MVPS, while 18% had either one or the other finding, and 10% had both (6).

We recently found that half of the 20 patients cardiologically evaluated in our own study sample had auscultatory and echocardiographic evidence of MVPS (7), a finding whose significance is as yet unclear. Panic patients with and without MVPS show no differences in family histories of panic disorder (8) or in response to imipramine (7). Gorman et al (7) speculate that while MVPS is not sufficient to account for panic attacks, it "may either play a provocative role" in predisposed individuals or "be linked to panic disorder through a generalized disturbance of the autonomic nervous system."

Pathophysiology

PRECIPITANTS OF PANIC DISORDER The first episode of panic attacks occurs in many patients, especially females, upon either threatened or actual separation from or loss of a loved person. Such patients often have histories of childhood separation anxiety as well. A common precipitant for male patients is a life-threatening event or illness. Another substantial proportion of patients seem to develop panic attacks in the context of endocrinological change. In one series of 32 patients, panic attacks began in three patients within a year of oophorectomy; in three within a year of hysterectomy; in three within one month postpartum; in two within a week prepartum; and in one within a week after beginning thyroid substitution treatment for hypothyroidism with associated amenorrhea (1). The patient reported above had a miscarriage and toxemia eight months before the onset of her panic attacks. Thus some phobic anxious patients appear to have a life-long vulnerability to separation-induced anxiety, while others have normal life adjustments until the onset of their anxiety symptoms following a traumatic life event or alteration of endocrine status (1).

One might be tempted to suggest at this point that in the case described above there exists a "clear-cut" psychodynamic explanation for the onset of the patient's panic attacks, namely that the difficulty with her boss reactivated the conflicted and unresolved childhood experience of the bullying father. However, psychodynamic formulations of agoraphobia, which led to psychoanalytically based treatment approaches, have in general proven rather unsuccessful (9, 10).

Nemiah (11) recently hypothesized that in patients with a biological vulnerability to panic, psychological conflict may give rise to panic attacks. In individuals without such a diathesis, similar conflicts would produce more general anxiety symptoms without discrete panic attacks. This conceptualization is compatible with the frequently reported finding of specific psychodynamic conflicts in some individuals with agoraphobia, without assigning etiological status in those conflicts. Such a "biological" formulation, implicit in Freud's early work on anxiety neurosis (12), was abandoned by later psychoanalytic investigations, which began to explain agoraphobia in exclusively psychological terms. This had the consequence of suggesting exclusively psychological treatment approaches, primarily psychoanalysis.

LACTATE INFUSIONS AND PANIC ATTACKS Despite the diversity of precipitating events, all patients with spontaneous panic attacks may have a common pathophysiological aberration, as suggested by the finding that intravenous lactate infusions will precipitate panic attacks in such individuals, but not in normal controls (13). As was first pointed out by Pitts & McClure (13), it is possible to regularly precipitate panic attacks, in patients prone to spontaneous episodes of panic, by the intravenous administration of sodium lactate. Furthermore, this occurs only rarely in normals, who usually experience rather mild symptomatology such as slight paresthesias and muscle tremor.

This finding, established under double-blind circumstances, has been amply replicated by other groups of investigators (14–16). It must be considered one of the better demonstrated facts in the experimental precipitation of psychopathological states. The meaning of this finding, however, still remains obscure. For instance, Pitts & McClure believe that the panicogenic effect was due to the chelation of calcium; they demonstrated that adding calcium to the sodium lactate prevented the occurrence of panic attacks. Fink, Taylor & Volavka (14) replicated this finding. However, Grosz & Farmer (17) consider it unlikely that lactate works via chelation of the calcium ion. Also, Pitts & McClure had given intravenous EDTA to a group of panic disorder patients who actually experienced tetany but not panic attacks. This area remains unclear and requires further investigation.

Ackerman & Sachar (18) also raised the question whether there was truly a specific panicogenic effect from lactate. They questioned whether there was not some sort of conditioned response whereby patients experienced the lactate infusion as producing autonomic symptomatology reminiscent of the onset of the panic attack and were therefore thrown into a panic attack, whereas normals had not developed such a learned sequence.

Although possible, this hypothesis seems contradicted by two main bodies of fact. First, EDTA infusions administered by Pitts produced profound hypocalcemic symptoms but no anxiety attacks in either anxiety neurotics or controls. Such infusions should certainly produce sufficient autonomic disturbance to throw patients into a panic attack if all that was required was a nonspecific trigger. Second, lactate-induced panic attacks resemble the spontaneous panic attacks with regard to the blocking effects of antidepressants. Kelly's group (16) induced panic attacks in eight agoraphobic patients and then treated them with phenelzine. Five of the eight patients responded well to the medication and did not have panic attacks upon reinfusion with lactate; the remaining three patients did not respond well to phenelzine and did reexperience panic attacks upon reinfusion with lactate. More recently our research group has administered lactate infusions to 25 patients with Panic Disorder; 16 experienced panic attacks. After clinical recovery with imipramine, and while still on the drug, eight were reinfused; not one experienced a panic attack.

MECHANISMS OF OTHER PHOBIC CONDITIONS Less is known about phobic conditions that do not involve panic attacks. Some simple phobias can be explained on the basis of simple conditioning, for example the avoidance of large dogs after being bitten by one. Other such fears can be taught or modeled by parental figures. Still others have been interpreted as having symbolic or unconscious meaning, e.g. the fear of heights as due to an unconscious wish to commit suicide by jumping. Klein (19) has also noted that some cases of Panic Disorder are preceded by a long history of claustrophobia, which may be based on subclinical autonomic arousal and panic attacks being triggered by situations of confinement.

The Social Phobias usually involve issues of low self-esteem. Usually the individual believes that others will detect signs of his anxiety in the phobic situation, creating a vicious cycle in which the irrational fear generates anxiety that in fact impairs performance.

Pharmacotherapy

TRICYCLIC ANTIDEPRESSANTS The recognition that imipramine would block panic attacks (1, 20–22) represented a major breakthrough in the treatment of anxiety and phobic disorders. Until that time agoraphobic patients with panic attacks had often been thought of as having a fear of open spaces or other specific situations, with long out-patient psychotherapies devoted to an exploration of unconscious sexual and aggressive impulses; the phobically barred areas were seen as symbols of forbidden temptations. However, while patients given imipramine began to report a dramatic decrease or cessation of panic attacks, many still appeared reluctant to change their phobic behavior patterns. Thus, it appeared that persuasion, direction, support, or minor tranquilizers were often required once the panic attacks had been blocked to help patients extinguish their anticipatory anxiety, which was not ameliorated directly by the tricyclic antidepressants, and to reenter the phobically avoided situations.

Once it has been ascertained that a patient is suffering spontaneous panic attacks and other medical and psychiatric conditions are excluded, a trial of imipramine is warranted. Interestingly, Panic Disorder patients are sometimes extremely sensitive to the drug, and become overstimulated on very low dosages, with jitteriness, irritability, or unusual energy accompanied by difficulty falling or staying asleep. In one series of 111 patients, Zitrin et al (22) found that 18% of the sample reported such side-effects on dosages of 5–75 mg per day. Thus, the recommended starting dose of imipramine for patients with panic attacks should be as low as 25 mg/day. Treatment schedules are described in greater detail in a recent publication (23).

Accompanying this increased sensitivity to stimulatory side-effects is the fact that the panic attacks themselves often respond more quickly and to lower dosages of imipramine than are required to treat depressive disorders. Whereas depressed patients usually require 150–300 mg/day for a period of several weeks for a clinical response, patients with panic attacks often report immediate beneficial effects once the appropriate dosage is reached, which may be anywhere between 5 and 300 mg/day. Even those patients who show stimulant side-effects can most often be treated effectively by lowering the dose and then increasing very gradually, sometimes to a maximum tolerated level of 10 mg/day (22). Patients should be warned in advance that unpleasant stimulatory side-effects do not preclude effective treatment; rather they suggest that the individual will clinically respond to minute doses of the drug. Finally, the measurement of plasma levels of imipramine and desimipramine in two patients successfully treated for panic attacks with imipramine 25 mg/day indicates that much lower drug plasma levels are required in this disorder than for depression, since one patient had a combined plasma drug level of between 55 and 60 ng/ml while the other had a level between 15 and 20 ng/ml (24). Depressives are generally thought to require combined imipramine-desmethylimipramine plasma levels in excess of 180 ng/ml for clinical benefit from imipramine (25).

Treatment should also include an investigation to determine the range of situations in which panic attacks occur, the common aspects of these situations, and life events related to the onset or worsening of symptoms. Finally, support, persuasion, and direction, sometimes accompanied by anti-anxiety drugs such as benzodiazepines, may be necessary to help patients overcome anticipatory anxiety and enter avoided situations, where they can begin to experience the fact that they will no longer have the panic attacks. This sometimes requires several months of gradually expanding activity, and can be adversely affected by a patient's breaking through the medication and having an episode of panic.

The question of how long to maintain the medication once symptom remission has occurred has not yet been clearly determined. As a general rule based on clinical experience, one can begin to taper the medication when a patient has been without panic attacks for 6–12 months. However, the disorder is usually episodic, and the patients should be advised that symptoms could reoccur at some future time, in which case they may need to go back on medication. While controlled data on relapse is sparse, one study found that of 11 agoraphobic patients who at least moderately improved after 26 weeks of imipramine and supportive psychotherapy, 4 (36%) relapsed within one year after cessation of treatment (22).

MONOAMINE OXIDASE INHIBITORS Monoamine oxidase inhibitors are also effective in blocking panic attacks and hence in treating phobic anxiety (26–29). Sheehan et al (30) recently reported a study in which phenelzine (45 mg/day) was at least as effective, if not more so, than imipramine (150 mg/day) in treating agoraphobics with panic attacks. These findings should not be construed, however, to indicate that phenelzine is the drug of choice for agoraphobia. The dosages in this study were low; the study should be repeated using up to 300 mg of imipramine and 90 mg of phenelzine. Also, given imipramine's demonstrated effectiveness and the added risk of hypertensive reactions from monoamine oxidase inhibitors, phenelzine should be considered a second-line drug, and not employed until a vigorous trial with a tricyclic has been undertaken.

OTHER CLASSES OF DRUGS Propranolol has also been given in an uncontrolled clinical trial to patients with pathological panic states (31) and was found effective in treating acute pathological panic. However, brief trials of doses up to 20 mg four times daily did not alleviate chronic panic attacks associated with agoraphobia. Two such patients developed a phobia to the propranolol and discontinued it; one then had a prompt and complete remission with imipramine. In a second study of propranolol vs placebo in the behavioral treatment of agoraphobics employing repeated in vivo exposures to stressful situations, propranolol given in single 40 mg doses 90 minutes before exposure negatively affected clinical response (32). The propranolol group spent significantly less time than the placebo group traveling alone in the month after treatment, and had improved significantly less on a measure of general symptoms at three months. The adverse influence of propranolol on treatment outcome appeared to be mainly due to a waning of drug effect toward the end of each five-hour exposure session.

In a detailed and thoughtful review of 16 controlled or partially controlled studies of beta-adrenergic blockers in anxiety disorders, Cole et al (33) concluded that this class of drugs may have a weak anti-anxiety effect. Specifically, the drugs often appear more helpful with somatic than with psychic aspects of anxiety. This conclusion was also reached in a recent study by Noyes et al (34), who did a propranolol-placebo cross-over study in patients with panic disorder. As such, it is useful for treating the anxiety symptoms that accompany the fear of public speaking and other social phobias. However, no controlled study has systematically examined the specific efficacy of propranolol in blocking panic attacks.

As mentioned before, minor tranquilizers are helpful in treating the anticipatory anxiety of Panic Disorder patients, but do not block the panic attacks. Preliminary findings of a controlled study by McNair & Kahn (35) appear to support this clinical impression. Panic attacks also are refractory to all classes of antipsychotic drugs, which may even exacerbate the condition. Because these patients are often considered borderline, and seen as decompensating when in acute panic, antipsychotics are often administered with a worsening of symptoms, use of more drug, further worsening, etc, leading to a vicious cycle (1, 20).

We are beginning to investigate the effect of clonidine on panic attacks. Clonidine is now used as an antihypertensive agent. Primate studies have shown, however, that clonidine, acting via $\alpha2$-adrenergic receptors, blocks the effects of electrical stimulation and pharmacological activation of the locus ceruleus, a major noradrenergic nucleus involved in activity of the sympathetic nervous sytem (36). Since sympathetic discharge is thought to be associated with symptoms of anxiety, it seems reasonable to hypothesize that clonidine may block panic attacks by blocking sympathetic overactivity. Several studies now suggest that this is worth pursuing. Gold et al (37) found that clonidine blocked the physical manifestations of the opiate withdrawal syndrome, while Svensson et al (38) reported that the drug reduced anxiety in patients with anxiety neuroses. While no investigator has yet completed a formal study of the effects of clonidine on panic attacks, our own limited experience suggests that it may be useful for some patients.

Behavior Therapy

Behavioral treatments for phobia involve structured, programmed exposure, either in imagination or in vivo, to feared objects or situations. A variety of treatment techniques, most commonly systematic desensitization and flooding, have been applied to both Simple Phobia and Agoraphobia.

Systematic desensitization involves having deeply relaxed patients approach or imagine themselves approaching a feared object or situation. Feared objects are arranged in a hierarchy and presented in a stepwise fashion, allowing gradual exposure. This contrasts with the technique of flooding, which is based on the principle of immediate confrontation to the most feared stimuli until one's anxiety abates. The difference can be likened to having someone who can swim but fears the water enter a pool from the shallow end and gradually work his way to the deep end, versus having him jump in at the deep end. These behavioral treatments have been administered in imagination by asking patients to visualize situations or presenting them with visual representations, and in vivo by having patients actually confront a feared situation or object.

Many investigations have shown systematic desensitization to be effective for treating circumscribed or simple phobias, but to be less effective for agoraphobia (39, 40). These data fit our hypothesis that behavior therapy is most useful for helping patients overcome anticipatory anxiety, which is the main reinforcer of phobic avoidance in the simple phobias. This would also apply to those cases of agoraphobia not accompanied by a history of panic attacks, as well as those in which panic attacks occurred at the onset of the illness but had since ceased to occur. However, we would expect that agoraphobics still having panic attacks would find systematic desensitization alone to be much less useful. In such cases a combination of drug treatment, to block panic attacks, and behavioral therapy, to overcome anticipatory anxiety and phobic avoidance, should be more effective than either treatment administered alone.

These hypotheses were tested by Zitrin et al (22, 41), who conducted a 26-week study of behavior therapy plus imipramine, behavior therapy plus placebo, and supportive psychotherapy plus imipramine in 218 patients classified as agoraphobics, simple phobics, or mixed phobics. The supportive psychotherapy was dynamically oriented and nondirective, while the behavior therapy consisted of relaxation training, systematic desensitization in imagination, specific in vivo desensitization homework assignments, and assertiveness training. The agoraphobics all suffered panic attacks, while the simple phobics had discrete phobias such as claustrophobia, fear of animals, social phobias, etc. The mixed phobics had spontaneous panic attacks and circumscribed phobias, but no travel restrictions.

Summarizing the patients' ratings of global improvement, agoraphobics and mixed phobics who received imipramine did significantly better than those who received placebo. In addition, agoraphobics and mixed phobics who received behavior therapy plus imipramine did significantly better than those who received behavior therapy plus placebo. Neither of these contrasts were significant for the simple phobics group, who did equally well on all these treatment regimens. Thus systematic desensitization without concomitant drug therapy seems maximally effective in phobics who are not experiencing ongoing panic attacks.

Behavioral investigators have also treated phobics with flooding in imagination and in vivo. Barrett (42) found that flooding in imagination worked more quickly than imaginal desensitization with snake phobias; while Strahley (43) reported the same results for flooding in vivo. In a prospective cross-over study of nine agoraphobics and seven "specific" phobics, Marks et al (44) reported that imaginal flooding (using frightening scenes) was significantly superior to imaginal desensitization, according to doctors ratings and physiological measures. Patients self-ratings, however, showed no significant difference between treatment in relation to main phobia, total phobia, and anxiety.

A third type of behavioral treatment falling midway between systematic desensitization and flooding has been termed "exposure in vivo." Successfully utilized with agoraphobics by Hand et al (45), patients in groups were exposed to what were for them increasingly difficult situations for four-hour periods on three days of one week. Zitrin et al (46) studied 76 agoraphobic women with combined group exposure in vivo and imipramine or placebo in a randomized double-blind study. A majority of the patients in both the placebo and imipramine groups showed moderate to marked improvement. However, group exposure plus imipramine was significantly superior to group exposure plus placebo on three of the four reported measures of improvement, including primary phobia, spontaneous panic, and global improvement. These data, which could not be accounted for by positing an antidepressant effect for the active drug, suggest that imipramine's antipanic effects augment the efficacy of group in vivo exposure for agoraphobia.

Case Report (continued)

Before concluding we wish to provide a follow-up on the patient described earlier. She was started on imipramine (25 mg HS) and encouraged to enter into situations where panic attacks had occurred. After two weeks of treatment she was able to go to restaurants without difficulty, but had a panic attack in a grocery store. The dosage was gradually increased to 100 mg qhs, at which time the patient was able to go to restaurants, supermarkets, and department stores, and also ride buses. Within several months she had taken a highly responsible job and was feeling more assertive and confident than she had in many months. However, she still found herself feeling close to panic at times in business meetings or on elevators, and imipramine was increased to 150 and then 200 mg qhs. On this dose she was able to function well for several months in all the situations in which she previously had become panicky.

After 10 months of treatment, she was passed over for a promotion and then criticized by her boss. She then had two panic attacks while walking alone on the street. Imipramine was increased to 250 mg qhs. Her sessions at this time were scheduled at three week intervals and were mainly supportive, although the patient's need for approval from, and difficulty both experiencing and expressing anger toward, male authority figures were explored. With this regimen she reported herself free of panic attacks for four months, and was able to function well professionally, while continuing her satisfying social and marital relationships. She also noted in retrospect that during the months she had been symptomatic she always felt more vulnerable to panic attacks during that part of her menstrual cycle between ovulation and menstruation.

Summary

Recent advances in our understanding of and ability to treat panic disorders and various types of phobia have been reviewed. A case of agoraphobia was presented in detail to illustrate clinical issues and serve as a basis for discussing differential diagnosis. The presence or absence of spontaneous panic attacks has crucial implications for classifying phobic states. Data and speculations concerning the pathophysiology of agoraphobia, social phobia and simple phobia were also presented. Finally, a variety of pharmacological and behavioral treatment approaches, some solidly established, others still investigational, were discussed.

Literature Cited

1. Klein, D. F. 1964. Delineation of two drug-responsive anxiety syndromes. *Psychopharmacologia* 5:397–408
2. Roth, M. 1959. The phobic anxiety-depersonalization syndrome. *Proc. R. Soc. Med.* 52:587–95
3. Pariser, S. F., Jones, B. A., Pinta, E. R., Young, E. A., Fontana, M. E. 1979. Panic attacks: diagnostic evaluations of 17 patients. *Am. J. Psychiatry* 136: 105–6
4. Frolich, E. D., Tarazi, R. C., Dustan, H. P. 1969. Hyperdynamic beta-adrenergic circulatory state. *Arch. Intern. Med.* 123:1–7
5. Easton, J. D., Sherman, D. G. 1976. Somatic anxiety attacks and propranolol. *Arch. Neurol.* 33:689–91
6. Markiewicz, W., Stoner, J., London, E. 1976. Mitral valve prolapse in one-hundred presumably healthy young females. *Circulation* 53:464–73
7. Gorman, J. M., Fyer, A. J., Gliklich, J., King, D., Klein, D. F. 1980. Mitral valve prolapse and panic disorders. effects of imipramine. In *Anxiety Revisited.* ed. D. F. Klein, J. G. Rabkin. Raven In press
8. Crowe, R. R., Pauls, D. L., Slymen, D. J., Noyes, R. 1980. A family study of anxiety neurosis. *Arch. Gen. Psychiatry* 37:77–79
9. Weiss, E. 1964. *Agoraphobia in the Light of Ego Psychology.* New York: Grune & Stratton
10. Klein, D. F., Liebowitz, M. R. 1980. *Psychotherapeutic attitudes in the drug treatment of anxiety and phobic reactions.* In *The Interrelationship of Psychotherapy and Psychopharmacology,* ed. M. H. Greenhill, A. Gralnick. New York: McMillan. In press
11. Nemiah, J. 1980. *The psychoanalytic view of phobias.* Presented at Ann. Meet. Am. Psychiatr. Assoc., 133rd, San Francisco, Calif.
12. Freud, S. 1955. A reply to criticisms of my paper on anxiety neurosis. In *The Standard Edition of the Complete Psy-*

chological Works of Sigmund Freud, 3:123–39. London: Hogarth

13. Pitts, F. N. Jr., McClure, J. N. Jr. 1967. Lactate metabolism in anxiety neuroses. *N. Engl. J. Med.* 277:1328–36
14. Fink, M., Taylor, M. A., Volavka, J. 1970. Anxiety precipitated by lactate. *N. Engl. J. Med.* 281:1129
15. Bonn, J. A., Harrison, J., Rees, W. L. 1971. Lactate-induced anxiety: therapeutic applications. *Br. J. Psychiatry* 119:468–71
16. Kelly, D., Mitchell-Heggs, N., Sherman, D. 1971. Anxiety and the effects of sodium lactate assessed clinically and physiologically. *Br. J. Psychiatry* 119: 129–41
17. Grosz, H. G., Farmer, B. B. 1969. Blood lactate in the development of anxiety symptoms: a critical examination of Pitts and McClure's hypothesis and experimental study. *Arch. Gen. Psychiatry* 21:611–19
18. Ackerman, S. M., Sachar, E. J. 1974. The lactate theory of anxiety: a review and re-evaluation. *Psychosom. Med.* 36(1):69–81
19. Klein, D. F., Gittelman, R., Quitkin, F. M., Rifkin, A. 1980. Diagnosis of anxiety, personality, somatoform, and factitious disorders. In *Diagnosis and Drug Treatment of Psychiatric Disorders: Adults and Children,* Chap. 15, pp. 493–521. Baltimore: Williams & Wilkins
20. Klein, D. F., Fink, M. 1962. Psychiatric reaction patterns to imipramine. *Am. J. Psychiatry* 119:432–38
21. Klein, D. F. 1967. The importance of psychiatric diagnosis in prediction of clinical drug effects. *Arch. Gen. Psychiatry* 16:118–26
22. Zitrin, C. M., Klein, D. F., Woerner, M. G. 1978. Behavior therapy, supportive psychotherapy, imipramine and phobias. *Arch. Gen. Psychiatry* 35: 307–16
23. Klein, D. F., Gittelman, R., Quitkin, F. M., Rifkin, A. 1980. Treatment of anxiety, personality, somatoform and factitious disorders. See Ref. 19, pp. 539–75
24. Jobson, K., Linnoila, M., Gillam, J., Sullivan, L. J. 1978. Successful treatment of severe anxiety attacks with tricyclic antidepressants: a potential mechanism of action. *Am. J. Psychiatry* 135:863–65
25. Glassman, A., Perel, J., Shostak, M., Kanotr, S., Fleiss, J. 1977. Clinical implications of imipramine plasma level in depressive illness. *Arch. Gen. Psychiatry* 34(3):197–204
26. Kelly, D., Guirguis, W., Frommer, E., Heggs, N. M., Sargant, W. 1970. Treatment of phobic states with antidepressants. *Br. J. Psychiatry* 116:387–98
27. Tyrer, P., Candy, J., Kelly, D. A. 1973. A study of the clinical effects of phenelzine and placebo in the treatment of phobic anxiety. *Psychopharmacologia* 32:237–54
28. Lipsedge, J. S., Hajjoff, J., Huggins, P., Napier, L., Pearce, J., Pike, D. F., Rich, M. 1973. The management of severe agoraphobia: a comparison of iproniazid and systematic desensitization. *Psychopharmacologia* 32:67–80
29. Solyom, L., Heseltine, G. F. D., McClure, D. J., Solyom, C., Ledgwidge, B., Steinburg, G. 1973. Behavior therapy versus drug therapy in the treatment of phobic neurosis. *Can. Psychiatr. Assoc. J.* 18:25–31
30. Sheehan, D. V., Ballenger, J., Jacobsen, G. 1980. Treatment of endogenous anxiety with phobic, hysterical and hypochondriacal symptoms. *Arch. Gen. Psychiatry* 37:51–59
31. Heiser, J. F., DeFrancisco, D. 1976. The treatment of pathological panic states with propranolol. *Am. J. Psychiatry* 133:1389–93
32. Hafner, J., Milton, F. 1977. The influence of propranolol on the exposure in vivo of agoraphobics. *Psychol. Med.* 7:419–25
33. Cole, J. O., Altesman, R. I., Weingarten, C. H. 1979. Beta-blocking drugs in psychiatry. *Psychopharmacol. Update, MacLean Hosp. J.*, pp. 40–68
34. Noyes, R., Kathol, R., Clancy, J., Crowe, R. 1980. Beta-blocking drugs: clinical evidence for site of anti-anxiety activity. See Ref. 7
35. McNair, D. M., Kahn, E. J. 1980. Imipramine vs. a benzodiazepine for agoraphobia. See Ref. 7
36. Redmond, D. E. 1979. New and old evidence for the involvement of a brain norepinephrine system in anxiety. In *Phenomenology and Treatment of Anxiety,* ed. W. E. Fann, I. Karacan, A. D. Pokorny, R. L. Williams, pp 153–203, New York: Spectrum
37. Gold, M. S., Redmond, D. E., Kleber, H. D. 1979. Noradrenergic hyperactivity in opiate withdrawal supported by clonidine reversal of opiate withdrawal. *Am. J. Psychiatry* 136:100–2
38. Svensson, T., Persson, R., Wallin, L., Walinder, J. 1978. Anxiolytic action of clonidine. *Nord. Psychiatr. Tidskr.* 32:439–41

39. Marks, I. M., Gelder, M. G. 1965. A controlled retrespective study of behaviour therapy in phobic patients. *Br. J. Psychiatry* 111:561–73
40. Gelder, M. G. Marks, I. M. 1966. Severe agoraphobia: a controlled prospective trial of behaviour therapy. *Br. J. Psychiatry* 112:309–19
41. Zitrin, C., Woerner, M. G., Klein, D. F. 1980. Distinction between anticipatory anxiety, panic anxiety, and avoidant behavior. See Ref. 7
42. Barrett, C. L. 1969. Systematic desensitization versus implosive therapy. *J. Abnorm. Psychol.* 74:587–92
43. Strahley, D. F. 1965. *Systematic desensitization and counterphobic treatment of an irrational fear of snakes.* Ph.D. thesis. Univ. Tenn. Chattanooga
44. Marks, I. M., Boulougouris, J., Marset, P. 1971. Flooding versus desensitization in the treatment of phobic patients: A crossover study. *Br. J. Psychiatry* 119:353–75
45. Hand, I., Lamontagne, Y., Marks, I. M. 1974. Group exposure (flooding) in vivo for agoraphobics. *Br. J. Psychiatry* 124:588–602
46. Zitrin, C. M., Klein, D. F., Woerner, M. G. 1980. Treatment of agoraphobia with group exposure in vivo and imipramine. *Arch. Gen Psychiatry* 37:63–72

Ann. Rev. Med. 1981. 32:601–13
Copyright © 1981 by Annual Reviews Inc. All rights reserved

CORONARY RISK FACTORS IN THE YOUNG

❖7421

Evan A. Stein, M.D., Charles J. Glueck, M.D., and John A. Morrison, Ph.D.

Department of Medicine, University of Cincinnati Medical Center,
Cincinnati, Ohio 45267

Almost two decades ago Holman et al (1) speculated that the genesis of the anatomic atherosclerotic lesion may be in childhood and may relate to excess intake of cholesterol and/or saturated fats. Subsequent studies support many of Holman's hypotheses. Coronary and aortic atherosclerosis was present in at least one third of teenage soldiers autopsied by Enos et al (2) and McNamara et al (3). Considerable anatomic evidence suggests that the adult atherosclerotic lesion (fibrous plaque and mature fibrocalcific plaque) evolves during the first two decades of life (1–3). Evidence that the ubiquitous fatty streak of childhood is a precursor of the mature atherosclerotic lesion has been proffered, but is by no means conclusive; it is necessarily based on postmortem data without longitudinal follow-up.

These and other studies have made it desirable to identify risk factors as early in life as possible, so that preventive measures can be introduced at a stage when they are most likely to influence the underlying pathologic process. The major established or suspected risk factors implicated in the development of atherosclerosis (elevated cholesterol, triglycerides, and blood pressure, early development of obesity, and smoking habits) can be recognized and dealt with in children.

Hypertension

Numerous epidemiological studies have implicated high blood pressure as a major risk factor associated with the development of atherosclerosis and its major vascular complications such as stroke and coronary artery disease. In the last decade there has been increasing awareness of, and interest in, hypertension as a pediatric problem. Since 1971 when Londe et al (4)

0066-4219/81/0401-0601$01.00

described hypertension in asymptomatic children, it has become generally accepted that essential hypertension, which affects 20–30% of American adults, has its origin and may be identified early in life (4–6).

A familial tendency toward hypertension has long been suspected by clinicians and is confirmed by population studies. A significant correlation of the blood pressures of siblings has been demonstrated (7). Studies on twins suggest a genetic explanation by revealing a closer correspondence of pressures among identical than among fraternal twins; the latter, in turn, correlate more closely than those of their siblings. However, environmental factors could explain the clustering of hypertension in families. Families share more than genes; even spouses tend to share blood pressures to some extent, and this in proportion to the duration of the marriage (8).

Parent-child correlation of blood pressures tends to be greatest in families in which spouse aggregation is also present. Furthermore, there has been no convincing demonstration of bimodality in the distribution of blood pressures. In 1971 Zinner et al (9) extended the findings of familial aggregation of blood pressure to include children 2–14 years old. Although this tends to favor a genetic explanation, early environmental influences of realtively short duration could also start the potential hypertensive patient on the road to overt disease. Proof of the environmental determinants of familial aggregation of blood pressure must still be awaited. However, their findings suggest that the process of essential hypertension, and ultimately its influence on the rate of atherogenesis, has its roots very early in life (6, 9). Most prospective studies of hypertension in adults also reveal that, despite its high prevalence, few entirely new cases evolve among normotensive persons after the age of thirty. Zinner et al (9) suggest that hypertension in adult life could possibly be predicted from the blood pressure pattern in youth. Direct evidence bearing on this is still awaited.

The predictive value of childhood blood pressure levels for adult hypertensive disease can be indirectly assessed by measuring the tracking correlation, that is, the difference between blood pressure measurements made in the same subjects at two different points in time. This technique has proven useful in adults (10). Even though an individual child or adolescent may have quite marked variability in his or her blood pressure on a day-by-day basis, in groups of children followed longitudinally tracking apparently does occur, at least from the age of six months onward (11, 12). The magnitude of the tracking correlations in early life are, however, much lower than those reported for adults.

With the availability of blood pressure distribution data for children and adolescents, better methods of measurement, and increased pediatrician awareness, increasing numbers of young persons are now being identified as hypertensive or at risk for developing hypertension in later life.

Blood pressure distribution curves for pediatric populations (10, 13) described by age, show a gradual increase in both systolic and diastolic values from infancy through adolescence. The Task Force on Blood Pressure Control in Children has proposed, therefore, that hypertension in the young should be defined as the presence of systolic and/or diastolic blood pressure levels persistently above the 95th percentile for any given year of life (13). The validity of the arbitrary cut-off point will be tested in the years ahead but has been complicated by the recent report by Voors et al (12) describing the influence of body size on blood pressure. From their studies it would appear that age has no direct relationship to blood pressure after adjusting for body mass. Considering the large variation in size of children at any given age, their proposed set of reference values for blood pressure that account for the child's height or weight is an interesting and new concept.

The relationship between elevated blood pressure in children and adolescents and obesity is well documented (14). It still remains to be determined whether this association is due to body mass per se, a relatively increased salt intake, altered salt handling, or a combination of these and other factors.

The benefit of early therapeutic intervention, even by dietary means, is likely to remain extremely difficult to quantitate since the natural history of primary hypertension detected in childhood is largely unknown. There are few or no studies upon which to base even a reasoned risk-to-benefit judgment with respect to early intervention. No reports exist, as yet, on either the efficacy of long-term therapy or the possible adverse complications associated with intervention. It is accepted (5) by a growing number of physicians that a persistently elevated diastolic blood pressure (> 95th percentile for age but < 100 mm Hg) in an otherwise asymptomatic child should be treated with weight loss if appropriate, combined with an exercise program and dietary salt reduction. While both weight reduction and salt restriction are still somewhat controversial as to their effectiveness in the treatment of essential hypertension in children, their use in the management of adult hypertensives has been well documented.

Lipids and Lipoproteins

The potency of total plasma cholesterol, low-density lipoprotein cholesterol (C-LDL), and high-density lipoprotein cholesterol (C-HDL) in the constellation of coronary heart disease risk factors dictates that careful attention be paid to levels of these lipids in children. Epidemiologic surveys of coronary heart disease (CHD) precursors in children indicate that differences in cholesterol and triglyceride levels appear early in childhood among cultures (15), among race-sex groups within cultures (16, 17), and among socioeconomic groups within cultures (16). Distributions of pediatric lipids in countries having significantly less CHD are shifted markedly toward lower total cholesterol values compared to the distributions in countries having high rates of CHD morbidity and mortality (15). Since cord blood cholesterol levels have been generally similar ($X = 65$ mg/dl) in all countries studied, the cross-cultural lipid-lipoprotein differences in older children are probably not "genetic" insofar as blood cholesterol levels reflect primarily "genetic" rather than "environmental" factors (18). Differences in lipid-lipoprotein levels appear by one year of age among countries and within countries between breast-fed and formula-fed infants.

Evidence of (clinically) significant changes in total lipid profiles during adolescence suggests that the origins of the sex-related differentials in CHD risk occur at this time (17). Prior to onset of sexual maturation, differences in total cholesterol, triglyceride, C-HDL, and C-LDL between sexes are minimal and not consistent (17).

Both cross-sectional and longitudinal evidences of a decrease in total cholesterol during adolescence have been reported. Based on a ten-year study of 35 boys and 28 girls, Lee (19) reported that 80% of the children experienced a decrease in total cholesterol during adolescence. The Cincinnati Lipid Research Clinic's (LRC's) Princeton prevalence study reported a significant cross-sectional decrease in total plasma cholesterol in white males and females and in black males, ages 12–16 years (17). There was also a significant increase in plasma triglyceride for all sex-race groups during the same period. These cross-sectionally observed changes had not come to full expression by age 14 years (17, 20). Cross-sectional analyses of lipoprotein data indicate that the decrement in total cholesterol occurring early in male adolescence is due primarily to a significant decrease (>12%) in the antiatherogenic C-HDL, and in white females, to moderate, statistically insignificant decreases in both C-HDL and C-LDL (Table 1) (17).

The significance of these adolescent changes in lipoptotein profiles is that the CHD-related sex-difference [males with higher C-LDL (atherogenic) and females with higher C-HDL (antiatherogenic)] appears at sexual maturity. If lipoprotein levels were to track over time, then high-risk lipoprotein profile adolescents would become high-risks adults. Data from the Cincinnati LRC's Princeton study (21) and from the studies in Muscatine (22) indicate that lipid and lipoprotein levels do track fairly well over time (4–6 years).

In a four-year follow-up, children who were 7–14 years old when they entered Cincinnati LRC's Princeton School District Study were found to have significant ($P<.001$) correlations between their plasma cholesterol

Table 1 Mean and (SD) cholesterol (TC), triglyceride (TG), high (C-HDL) and low (C-LDL) density lipoprotein cholesterol (MG/DL) in 927 fasting school children[a]

Age	*Male Subjects*										*Female Subjects*									
	White					*Black*					*White*					*Black*				
	No.	*TC*	*TG*	*C-HDL*	*C-LDL*	*No.*	*TC*	*TG*	*C-HDL*	*C-LDL*	*No.*	*TC*	*TG*	*C-HDL*	*C-LDL*	*No.*	*TC*	*TG*	*C-HDL*	*C-LDL*
6-7	50	158 (21)	50 (16)	56 (11)	94 (21)	17	158 (31)	49 (15)	60 (15)	93 (26)	39	161 (16)	71 (35)	49 (9)	101 (19)	16	168 (15)	55 (14)	57 (11)	104 (18)
8-9	45	158 (20)	57 (25)	56 (10)	92 (18)	18	172 (29)	48 (14)	62 (13)	103 (23)	40	162 (21)	70 (28)	53 (11)	98 (19)	21	176 (22)	56 (20)	62 (8)	104 (23)
10-11	60	164 (31)	58 (20)	56 (10)	98 (29)	20	171 (27)	58 (20)	63 (9)	99 (29)	70	161 (25)	72 (24)	51 (10)	98 (21)	22	160 (22)	69 (36)	53 (10)	99 (20)
12-13	78	160 (24)	65 (29)	55 (15)	96 (22)	28	167 (38)	59 (21)	55 (13)	102 (39)	56	161 (25)	73 (24)	52 (10)	98 (20)	32	168 (25)	69 (28)	56 (13)	102 (24)
14-15	57	155 (29)	69 (29)	48 (10)	96 (27)	23	156 (22)	60 (22)	53 (10)	95 (17)	56	151 (21)	72 (27)	50 (10)	91 (19)	21	166 (18)	60 (18)	56 (11)	102 (19)
16-17	68	155 (24)	78 (39)	48 (10)	94 (22)	13	145 (19)	63 (18)	50 (12)	86 (18)	65	159 (24)	72 (33)	55 (12)	93 (22)	12	152 (24)	62 (19)	54 (10)	89 (30)
6-11	155	160 (25)	55 (21)	56 (11)	95 (24)	55	167 (29)	52 (17)	62 (12)	98 (26)	149	161 (22)	71 (28)	51 (10)	99 (20)	59	168 (21)	60 (27)	58 (10)	102 (20)
12-17	203	157 (25)	70 (33)	51 (12)	96 (23)	64	159 (30)	60 (21)	54 (12)	97 (29)	177	157 (24)	72 (28)	52 (11)	94 (21)	65	164 (23)	65 (24)	56 (12)	100 (24)

[a] Reprinted from reference 17.

levels at year one and at years two (r=.65), three (r=.66), and four (r=.68) (21). Lower but significant correlations were observed for triglyceride (r = .40). Both C-HDL and C-LDL levels as subsequent examinations (three and four years later) were very significantly correlated with initial lipoprotein results; C-HDL had correlations of .60 and .53 and C-LDL of .67 and .61 at years 4 and 5. Similar correlations between total cholesterol determined at entrance and two, four, and six years later were found in the Muscatine study (22).

From these data on tracking, one is led to speculate that, on average, adolescents with a total cholesterol and C-LDL in the highest age-specific quintiles of these distributions will become adults with total cholesterol and C-LDL in the highest quintile of these atherogenic distributions. That is, adolescents with total cholesterol >180 and C-LDL >110 mg/dl may well have levels of 240 and 160 mg/dl respectively as adults.

The practicality of pediatric lipid and lipoprotein screening depends on case-finding efficiency, lability of lipids in children, and ability to intervene successfully. Case-finding efficiency in children is limited by the fact that cut-off points customarily used to separate "normal" from "abnormal" plasma lipid and lipoprotein levels are based on extreme percentiles (e.g. 95th or 90th percentile) that, by definition, make a case a relatively rare event. Since C-LDL and C-HDL have a continuous relationship to CHD events throughout their distribution, it is suggested that risk detection should not be tied so rigidly to arbitrary cut-off points. This same problem is faced in adult screening; establishment of adult cut-off points can be informed by other strategems such as that level at and above which intervention has demonstrably beneficial effects (23). Since children with elevated risk factors exhibit clinical disease only in the rarest of circumstances, and since longitudinal data demonstrating that children with elevated risk factors will be at increased risk to CHD as adults do not exist, such a stratagem is not practical. A body of data does exist, however, that suggests a link between pediatric risk factors and subsequent (possible) CHD, which can result in improved case-finding efficiency. These are the data on the familial nature of CHD and familial aggregation of CHD risk factors (24, 25).

First, children born to parents with early (< 50 years of age) myocardial infarction have significantly increased cholesterol levels relative to control children born to parents without early myocardial infarction (24, 25). In a reversed approach to the same problem, Schrott et al (26) compared cardiovascular histories of first-degree relatives of children with high (≥ 95th percentile), intermediate (5–95th percentile), and low (< 5th percentile) cholesterol. They reported significantly increased cardiovascular mortality in families of high cholesterol children. Second, data from numerous studies indicate that significant correlations for total cholesterol, triglyceride, C-HDL, and C-LDL exist between parents and their children (27–30).

Data from the Princeton study indicate that children born in and living in households with at least one high-cholesterol parent ($\geq$ age-, sex-, race-specific 95th percentile) were over twice as likely ($P < .01$) to have elevated cholesterol themselves ($\geq$ age-, sex-, race-specific 95th percentile) (30). Thus, children whose parents are found to have elevations of lipids and lipoproteins should also be tested.

The lability of pediatric risk factors raises the issue of the significance of a one-time measurement. For adults, the Framingham study has shown that a one-time measurement of total cholesterol, C-LDL, C-HDL, systolic, and diastolic blood pressure, severally and collectively do track and do predict future CHD (31). For children, there is a growing body of data, partially reviewed above, suggesting that blood lipids, blood pressures, and body mass measurements do track fairly well.

The effect of the phenomenon known as "regression toward the mean" on extreme values has been discusssed recently as legislating against pediatric screening (32). Briefly put, the "regression toward the mean" phenomenon means that, under conditions of repeated measurements, subjects with extreme values on one measurement will tend to have less extreme values, closer to the mean, on remeasurement (33). In the Princeton LRC's study, 42% (138 of 326) of children with total cholesterol $\geq$ 205 mg/dl (the 95th percentile) had values below that cut-off point at retest (33). This fact does not, however, argue against pediatric screening. The mean total cholesterol for these 138 children at retest was approximately the 90th percentile. Since the risk to CHD associated with total cholesterol is continuous, this cohort, as a whole, represents an elevated risk group. Repeat testing prior to therapeutic intervention is needed, however, to identify patients with sporadic or transient elevations and to evaluate fully and accurately the efficacy of the intervention.

Pediatric treatment of hyperlipidemia aimed at reducing elevated plasma lipid levels is ultimately directed to the primary prevention of atherosclerosis. Although familial and acquired hyperlipidemias are associated with accelerated coronary heart disease, there is still a lack of consistent, unequivocal evidence that treatment of hyperlipidemia, antecedent to a clinical or pathologic atherosclerotic event, will prevent or retard development of atherosclerosis.

Whether the dietary intervention is a moderate one used for all populations or tailored for the lipid elevation in limited populations of hyperlipidemic children, it should be scrutinized carefully for both utility and potential hazards of modifying traditional dietary practices. In evaluation and recommendations of diets for young children, most attention has been placed on the content of cholesterol and saturated, polyunsaturated, and total fats. A number of questions have been raised about the safety of low-cholesterol, high-polyunsaturated-fat, and high-phytosterol diets;

1. Relatively high cholesterol intake may be necessary in the first year of life to satisfactorily induce cholesterol metabolic systems that would otherwise be less than optimally functional later in life (34, 35). This concern is based on observations of male, but not female, rats and swine. When fed cholesterol-rich diets in infancy, these animals had lower serum cholesterol levels when fully grown than did the control group initially fed low-cholesterol diets. However, these provocative findings have not been replicated in four studies of normal and hypercholesterolemic infants and children (36–38).
2. Cholesterol is an essential nutrient in early infancy, and central and peripheral nervous system myelination might be affected by low-cholesterol intake at that time (39). However, almost all brain cholesterol is derived from endogenous biosynthesis (40), and it is unlikely that dietary cholesterol plays any important role relative to brain cholesterol. Moreover, Plotz et al (41) reported that in the brain of the human fetus, cholesterol is synthesized in situ from glucose.
3. Cholesterol-lowering diets might lead to overall deficiencies of essential nutrients. As pointed out by Fomon (42) in 1974, cholesterol-modified diets would not restrict lean meat, fish, poultry, vegetable sources of protein, or skim milk. However, on strict low-cholesterol diets in two-year-old or younger children, iron requirements may not be met without some supplemental iron provision (43).
4. Low-cholesterol, vegetable-oil-rich, phytosterol-rich diets might elevate plasma and tissue phytosterol levels in infancy and childhood. In children on cholesterol-lowering diets and in infants receiving vegetable-oil-rich, standard commercial formula, three- to five-fold increments in plasma phytosterols have been reported (44). In the first several months of life, vegetable oil formula-fed infants accrue plant sterols in their aortic tissue, although quantities are in micrograms as compared to milligrams for cholesterol. Much larger amounts of phytosterols are present in mature atheromatous lesions in adult. The relationship of tissue phytosterol to the deposition of tissue cholesterol in atheromatous and normal aortic tissue remains to be elucidated.
5. There may be an enhanced incidence of gallstones on polyunsaturate-rich, low-cholesterol diets, a finding in adults by Sturdevant et al (45). However, Meittinen et al (46) failed to find any excess of gallstones in a carefully controlled study of subjects receiving a cholesterol-poor, polyunsaturate- and phytosterol-rich diet.

6. Diets rich in polyunsaturated fats have been suggested to be related to colonic cancer. The subject has been reviewed (47, 48), and both reports concluded that this idea is not supported by clinical experience, epidemiological research, or experimental studies.

There is a vague and unspecified uneasiness that certain proposed dietary changes might be followed by currently unanticipated adverse consequences. The utility of making dietary changes must be balanced against this anxiety and against the supposition that diet would have significant beneficial effects. At the current state of the art, it appears that the benefits will be likely to overshadow both the known and unanticipated risks.

Smoking, Drinking, and Oral Contraceptives

Epidemiologic studies of environmental factors that affect C-HDL levels have identified obesity (12, 20) cigarette smoking (49), alcohol consumption (50), and use of estrogen-progestin oral contraceptives as having substantive effects on adult levels of C-HDL (49).

The Cincinnati LRC's Princeton prevalence program studied the relation of smoking, alcohol consumption, and oral contraceptive usage to lipids and lipoproteins in school children, ages 12–19 years (51). After pair-matching for age, sex, race, and total cholesterol, adolescent children who reported smoking cigarettes (mean number of cigarettes daily was eight) had significantly lower C-HDL (–6.1 mg/dl) and higher C-LDL (+4.1 mg/dl) than nonsmokers. The mean decrement in C-HDL (–6.1 mg/dl) is well within the difference in C-HDL between CHD patients and controls of 4–7 mg/dl observed by Castelli et al (52). The smoking habit is usually acquired early in life, especially in adolescents who are exposed to it by their peers, siblings, and parents. Although it is unlikely that the cigarette habit represents a serious threat to cardiovascular health in adolescence, the toll is exacted later in life after the habit becomes established and the cardiovascular system is more vulnerable to its effects.

The differences in C-HDL in adolescents who reported drinking alcoholic beverages twice a week or more frequently were in the expected direction (drinkers higher) but were not statistically significant (51). The frequency of drinking (twice a week) and quantity of alcohol consumed by this adolescent cohort (51) were considerably less than by adult men in the study of Castelli (50).

Wallace et al (53) studied lipid-lipoprotein levels in 31 adolescents users of oral contraceptives and controls from the collaborative LRC study matched for age, race, and relative ponderosity. Adolescent females using oral contraceptives had significantly higher C-HDL and triglyceride, and modestly higher C-LDL (53). No differences in the Quetelet index (weight/height^2) were found between users and nonusers (53).

Body Mass

The health hazard of being overweight has been appreciated for hundreds, if not thousands, of years. That fat people die earlier was emphasized by Hippocrates, and later Maimonides and others. However, some of the major epidemiological studies of CHD have not demonstrated an independent relationship between CHD and gross obesity (54). This is all the more unexpected since obesity and hypertension are correlated. Despite earlier evidence to the contrary, recent studies have shown a positive association with systolic and diastolic blood pressure, triglyceride and low-density lipoprotein cholesterol, while high-density lipoprotein cholesterol is negatively associated with ponderosity as measured by the Quetelet index or skinfold thickness (12, 20).

The link between obesity in infancy and the condition in adulthood is inferred from a number of separate studies. Several connect overweight in infancy to obesity at the age of 5–7 years, and others correlate mid-childhood and early-adult obesity (55–57). Although no single study has followed children from birth to adulthood, the inference for these separate studies is suggestive. These studies raise significant questions on the role of genetics vs nutritional excess. Is the weight attained a causal factor itself or merely an expression of some genetic predisposition? Several studies speculate that only a proportion of weight gain in infancy can be attributed to variations in dietary intake (56). It is still speculative if limiting weight gain in childhood will have any long-term influence on adult weight status. These questions cannot be answered with information now available. But the present data do indicate that it is possible to identify rather early a population at risk of becoming overweight adults. Whether that can or should lead to intervention remains to be clarified.

Summary

The exceptional growth in risk factor assessment and the efficacy of primary prevention in childhood of atherosclerotic and hypertensive diseases is the subject of entire books that have recently collated the results of multiple investigators (58, 59). It is a topic well beyond the constraints of this review. It appears that the major "adult" CHD risk factors, including elevated total plasma and low-density lipoprotein cholesterol, elevated triglyceride, reduced high-density lipoprotein cholesterol, high blood pressure, obesity, and initiation of cigarette smoking can usually be recognized in children. There are, as yet, no longitudinal studies that control for one or more of these CHD risk factor variables to determine whether future development of CHD could be prevented or ameliorated. In the absence of unequivocal longitudinal studies of efficacy of intervention, prudent, safe, and well-supervised interventions should be carried out only after exhaustive proof of diagnosis (58). The recently summarized data suggest that the hyperlipoproteinemias, high blood pressure, obesity, and initiation of cigarette smoking can (with varying degrees of success) be dealt with during childhood and adolescence.

Literature Cited

1. Holman, R. L. 1961. Atherosclerosis—a pediatric nutrition problem? *Am. J. Clin. Nutr.* 9:565–69
2. Enos, W. F., Beyer, J. C., Holmes, R. H. 1955. Pathogenesis of coronary disease in American soldiers killed in Korea. *J. Am. Med. Assoc.* 158:912–14
3. McNamara, J. J., Molot, M. A., Stremple, J. F., Cutting, R. T. 1971. Coronary artery disease in combat casualties in Vietnam. *J. Am. Med. Assoc.* 216:1185–87
4. Londe, S., Bourgoingnie, J. J., Robsin, A. M., Goldring, D. 1971. Hypertension in apparently normal children. *J. Pediatr.* 78:569
5. Loggie, J. M. H., New, M. I., Robson, A. M. 1979. Hypertension in the pediatric patient: a reappraisal. *J. Pediatr.* 94:685–99
6. Levine, L. S., Hennekins, C. H., Klein, B., Gourley, J. E., Briese, F. W., Hokanson, J., Gelband, H., Jessee, M. J. 1978. Tracking correlations of blood pressure during infancy. *Pediatrics* 61:121
7. Feinbeib, M., Garrison, R., Borhani, N., et al. 1975. Studies of hypertension in twins. In *Epidemiology and Control of Hypertension,* ed. O. Paul, 1:3–17. New York: Stratton Intercontinental Medical Book Corp.
8. Winkelstein, W. Jr., Kantor, S., Ibrahim, M. 1966. Familial aggregation of blood pressure. *J. Am. Med. Assoc.* 195:858–90
9. Zinner, S. H., Levy, P. S., Kass, E. H. 1971. Familial aggregation of blood pressure in childhood. *N. Engl. J. Med.* 284:401
10. Lauer, R. M., Connor, W. E., Leaverton, P. E., Reiter, M. A., Clarke, M. S. 1975. Coronary heart disease risk factors in school children: the Muscatine Study. *J. Pediatr.* 86:697
11. Berenson, G. S., Blonde, C. V., Farris, R. P., Foster, T. A., Frank, G. C., Srinivasan, S. R., Voors, A. W., Webber, L. S. 1979. Cardiovascular disease risk factor variables during the first year of life. *Am. J. Dis. Child.* 133:1049–58
12. Voors, A. W., Webber, L. S., Frerichs, R. R., Berenson, G. S. 1977. Body height and body mass as determinants of basal blood pressure in children—the Bogalusa Heart Study. *Am. J. Epidemiol.* 106:101
13. Report of the Task Force on Blood Pressure Control in Children. 1977. *Pediatrics* 59(Suppl):797
14. Levine, L. S., Lewy, J. E., New, M. I., 1976. Hypertension in high school students, evaluation in New York City. *NY State J. Med.* 76:40
15. Golubjatnikov, R. T., Paskey, T., Inhorn, S. L. 1972. Serum cholesterol levels of Mexican and Wisconsin school children. *Am. J. Epidem.* 96:36–39
16. Srinivasan, S. R., Frerichs, R. R., Webber, L. S., et al 1976. Serum lipoprotein profile in children from a biracial community. The Bogalusa Heart Study. *Circulation* 54:309–18
17. Morrison, J. A., deGroot, I., Edwards, B. K., et al 1978. Lipids and lipoproteins in 927 school children, ages 6–17 years. *Pediatrics* 62(6):990–95
18. Klimov, A. N., Glueck, C. J., Gartside, P. S., et al 1979. Cord blood high density lipoproteins: Leningrad and Cincinnati. *Pediatr. Res.* 13:208–10
19. Lee, V. A. 1967. Individual trends in the total serum cholesterol of children and adolescents over a ten-year period. *Am. J. Clin. Nutr.* 20:5–12
20. Laskarzewski, P. M., Morrison, J. A., Mellies, M. J., et al. 1980. Relationships of measurements of body mass to plasma lipoproteins in school children and adults. *Am. J. Epidemiol.* 111:395–407
21. Laskarzewski, P., Morrison, J. A., deGroot, I., et al 1979. Lipid and lipoprotein tracking in 108 children over a four-year period. *Pediatrics* 64:584–91
22. Clarke, W. R., Schrott, H. G., Leaverton, P. E., et al. 1978. Tracking of blood lipids and blood pressures in school-age

children: the Muscatine Study. *Circulation* 58:626–34

23. Veterans Administrative Cooperative Study Group on Antihypertensive Agents: Effects of Treatment on Morbidity in Hypertension: Parts I and II. 1970. *J. Am. Med. Assoc.* 213(7):1028–34, 1043–52
24. Glueck, C. J., Fallat, R. W., Tsang, R., et al. 1974. Hyperlipidemia in progeny of parents with myocardial infarction before age 50. *Am. J. Dis. Child.* 127:70–75
25. Hennekens, C. H., Jesse, M. J., Klein, B. E., et al. 1976. Cholesterol among children of men with myocardial infarction. *Pediatrics* 58(2):211–17
26. Schrott, H. G., Clark, W. R., Wiebe, D. A., et al. 1979. Increased coronary mortality in relatives of hypercholesterolemic school children—the Muscatine Study. *Circulation* 59:320–26
27. Hewitt, D., Jones, G. J. L., Godin, G. J., et al. 1979. Nature of the familial influence on plasma-lipid levels. *Atherosclerosis* 32(4):381–96
28. Morrison, J. A., Khoury, P., Samuels, S. J., et al. 1980. Intrafamilial associations of cholesterol and triglyceride among related and unrelated household members. *Clin. Genet.* In press
29. Garrison, R. J., Castelli, W. P., Feinleib, M., et al. 1979. The association of total cholesterol, triglycerides and plasma lipoprotein cholesterol levels in first degree relatives and spouse pairs. *Am. J. Epidemiol.* 110(3):313–21
30. Morrison, J. A., Kelly, K. A., Mellies, M. J., et al. 1978. Parent-child associations at upper and lower ranges of plasma cholesterol and triglyceride. *Pediatrics* 62:468–78
31. Kannel, W. B., Castelli, W. P., Gordon, T., et al. 1971. Serum cholesterol, lipoproteins and the risk of coronary heart disease. *Ann. Intern. Med.* 74:1–12
32. Holtzman, N. A. 1979. Hyperlipidemia screening—search for heffalumps (Editorial). *Pediatrics* 64(2):270–71
33. Morrison, J. A., Laskarzewski, P., deGroot, I., et al. 1979. Diagnostic ramifications of repeated plasma cholesterol and triglyceride measurements in children: regression towards the mean in a pediatric population. *Pediatrics* 64(2):197–201
34. Fomon, S. 1971. A pediatrician looks at early nutrition. *Bull. NY Acad. Med.* 47:569–78
35. Hahn, P., Kirby, L. 1973. Immediate and late effects of premature weaning and of feeding a high fat or high carbohydrate diet to weanling rats. *J. Nutr.* 103:690–96
36. Glueck, C. J., Tsang, R., Balistreri, W., Fallat, R. 1972. Plasma and dietary cholesterol in infancy: effects of early low or moderate dietary cholesterol intake on subsequent response to increased dietary cholesterol. *Metabolism* 21:1181–92
37. Friedman, G., Goldberg, S. J. 1975. An evaluation of the safety of a low saturated diet, low cholesterol diet, beginning in infancy. *Pediatrics* 58(5):655–57
38. Hodgson, P. A., Effefson, R. D., Elveback, L. R., Harris, L. E., Nelson, R. A., Weidman, W. H. 1976. Comparison of serum cholesterol in children fed high, moderate, or low cholesterol milk diets during neonatal period. *Metabolism* 25(7):739–46
39. Davison, A. N. 1973. Biology of normal brain development. *Pediatr. Res.* 7:48
40. Morris, M. D., Chaikoff, I. L. 1961. Concerning incorporation of labelled cholesterol, fed to mothers, into brain cholesterol of 20-day-old suckling rats. *J. Neurochem.* 8:226–29
41. Plotz, E. J., Kabara, J. J., Davis, M. E., LeRoy, G. V., Gould, R. G. 1968. Studies on the synthesis of cholesterol in the brain of the human fetus. *Am. J. Obstet. Gynecol.* 101:534–38
42. Fomon, S. J. 1974. *Infant Nutrition.* Philadelphia: Saunders. 2nd ed.
43. Larsen, R., Glueck, C. J., Tsang, R. C. 1974. Special diet for familial Type II hyperlipoproteinemia. *Am. J. Dis. Child.* 128:67–72
44. Mellies, M., Glueck, C. J., Sweeney, C., Fallat, R. W., Tsang, R., C., Ishikawa T. T. 1976. Plasma and dietary phytosterols in children. *Pediatrics* 57(1):60–67
45. Studevant, R. A. L., Pearce, M. L., Dayton, S. 1973. Increased prevalence of cholelithiasis in men ingesting a serum cholesterol lowering diet. *N. Engl. J. Med.* 228:24–27
46. Miettinen, O., Turpeinen, M., Karvonen, J., Paavilainen, E., Elosuo, R. 1976. Prevalence of cholelithiasis in men and women ingesting a serum cholesterol lowering diet. *Ann. Clin. Res.* 8:111–16
47. Wynder, E. L., Reddy, B. S. 1975. Dietary fat and colon cancer. *J. Natl. Cancer Inst.* 54:7–10
48. Heyden, S. 1974. Polyunsaturated fatty-acids and colon cancer. *Nutr. Metab.* 17(6):321–28
49. VanGent, G. M., VanderVoot, H., Hessel, L. W. 1978. High density lipo-

protein cholesterol, monthly variation, and association with cardiovascular risk factors in 1,000 forty-year-old Dutch citizens. *Clin. Chim. Acta* 88:155–62

50. Castelli, W. P., Doyle, J. T., Gordon, T., et al. 1977. Alcohol and blood lipids: the cooperative lipoprotein phenotyping study. *Lancet* 2:153–55
51. Morrison, J. A., Kelly, K. A., Mellies, M. et al. 1979. Cigarette smoking, alcohol intake, and oral contraceptives: relationships to lipids and lipoproteins in adolescent school children. *Metabolism* 28:1166–70
52. Castelli, W. P., Doyle, J. T., Gordon, T., et al. 1977. HDL cholesterol and other lipids in coronary heart disease. *Circulation* 55:767–72
53. Wallace, R. B., Tamir, I., Heiss, G., et al. 1979. Plasma lipids, lipoproteins, and blood pressure in female adolescents using oral contraceptives. *J. Pediatr.* 95:1055–59
54. Dawber, T. R., Moore, F. E., Mann, G. V. 1957. *Am. J. Publ. Health* 47(suppl. 4):1–63
55. Asher, P. 1966. Fat babies and fat children. *Arch. Dis. Child.* 41:672–73
56. Mellbin, T., Vuille, J. C. 1973. Physical development at seven years of age in relation to velocity of weight gain in infancy with special reference to incidence of overweight. *Br. J. Prev. Soc. Med.* 27:225–35
57. Fisch, R. O., Bilek, M. K., Ulstrom, R. 1975. Obesity and leanness at birth and their relation to body habitus in later childhood. *Pediatrics* 56:421–28
58. Lauer, R., Shekelle, R. B., eds. 1980. *Childhood Prevention of Atherosclerosis and Hypertension.* New York: Raven
59. Strong, W. B. 1978. *Atherosclerosis: Its Pediatric Aspects.* New York: Grune & Stratton

CUMULATIVE INDEXES

CONTRIBUTING AUTHORS, VOLUMES 28–32

A

Abildskov, J. A., 30:171–79
Ahmed, S. S., 28:393–409
Ahn, Y. S., 28:299–309
Alfrey, A. C., 29:93–98
Annas, G. J., 29:9–14
Anthony, B. F., 28:355–69
Archer, D. B., 32:101–6
Arnold, A. B., 28:187–94
Arnon, S. S., 31:541–60
Athanasoulis, C. A., 30:41–59
Atkinson, J. P., 28:291–98

B

Babior, B. M., 32:313–26
Bailey, J. C., 29:417–26
Barnes, R. W., 31:201–18
Begemann, H., 28:131–41
Behar, J., 32:443–56
Bergsagel, D. E., 30:431–43
Betts, R. F., 28:103–10
Bierman, E. L., 30:155–70
Bigger, J. T. Jr., 30:91–118
Blankenhorn, D. H., 30:289–300
Bollet, A. J., 31:463–87
Bonadonna, G., 29:149–75
Bond, J. H., 31:127–37
Bone, H. G. III, 28:111–17
Bouchier, I. A. D., 31:59–77
Boyar, R. M., 29:509–20
Branch, R. A., 29:205–18
Braylan, R. C., 30:17–24
Brennan, M. F., 32:233–43
Brenner, B. M., 28:91–102
Bresnick, G. H., 30:331–38
Bressler, R., 31:329–43
Bristow, J. D., 29:475–83
Broadus, A. E., 32:457–73
Burke, J. S., 30:17–24
Buselmeier, T. J., 31:345–60

C

Cantor, H., 30:269–77
Caro, W. A., 28:251–69
Cavalieri, R. R., 28:57–65
Cerami, A., 31:29–34
Chandra, N., 32:435–42
Charache, S., 32:195–206
Cicero, T. J., 32:123–42
Cobb, L. A., 31:453–62
Coceani, F., 32:375–85
Cohen, J. J., 29:521–41
Coleman, T. G., 31:15–27
Colman, N., 31:433–39
Colman, R. W., 30:359–74
Colvin, R. B., 29:191–203
Conklin, R. H., 28:167–77
Cooper, T. B., 30:259–68
Copass, M. K., 31:453–62
Crawford, D. W., 30:289–300
Crosnier, J., 29:67–72

D

Dallman, P. R., 32:143–60
D'Angio, G. J., 30:405–15
Daniels, J. R., 32:543–57
DeClue, J. W., 31:15–27
Douglas, R. G. Jr., 31:233–38
Droegemueller, W., 31:329–43
Dudrick, S. J., 28:517–28
DuPont, H. L., 28:167–77
Dustan, H. P., 29:485–93
Dwyer, J. M., 32:491–97

E

Ehrhardt, A. A., 30:417–30
Eickhoff, T. C., 28:411–23
Elharrar, V., 29:417–26
Ellis, A. K., 31:489–508
Ensinck, J. W., 30:155–70
Epstein, C. J., 29:117–28
Ettinger, P. O., 28:393–409
Ettlinger, R. E., 29:15–22
Evenson, M. A., 28:85–90
Everett, R. B., 30:473–88

F

Fajans, S. S., 30:313–29
Fanelli, G. M. Jr., 28:349–54
Fang, L. S. T., 30:225–39
Feinstone, S. M., 29:359–66
Fekety, R., 32:327–33
Fernstrom, J. D., 32:413–25
Fialkow, P. J., 30:135–43
Finegold, S. M., 31:219–32
Fink, M., 32:405–12
Fishman, A. P., 31:421–31
Fitzgerald, F., 29:177–89, 32:221–31
Floyd, J. C. Jr., 30:313–29
Foreman, J. C., 31:181–90
Forstot, S. L., 28:21–35
Frank, M. M., 28:291–98
Franklin, E. C., 32:261–71
Freeman, H. J., 29:99–116
Freis, E. D., 30:81–89
Friedman, E. A., 31:463–87
Froelicher, V. F. Jr., 28:1–12

G

Gabig, T. G., 32:313–26
Gasparini, M., 29:149–75
Gennari, F. J., 29:521–41
Gerald, P. S., 29:563–70
Giannella, R. A., 32:341–57
Giebink, G. S., 29:285–306
Gill, J. R. Jr., 31:405–19
Gillis, R. A., 29:73–79
Ginzler, E. M., 31:463–87
Gitlin, N., 32:161–74
Glatstein, E., 29:81–91
Glueck, C. J., 32:601–13
Golbus, M. S., 29:117–28
Goldberg, H. I., 30:181–88
Goldfine, I. D., 29:407–15
Goodnight, J. E. Jr., 29:231–83
Goodwin, D. W., 32:93–99
Gorevic, P. D., 32:261–71
Graber, S. E., 29:51–66
Greer, M. A., 31:385–404
Grossman, M., 28:43–55
Grossman, M., 32:45–61
Grossman, M. I., 29:495–507
Grumet, F. C., 28:425–52
Gutterman, J. U., 28:489–515
Guyton, A. C., 31:15–27

H

Habener, J. F., 29:327–42
Haider, B., 28:393–409
Hall, C. L., 30:199–211
Hall, W. J., 31:233–38
Hallstrom, A. P., 31:453–52
Halmi, K. A., 29:137–48
Halsted, C. H., 31:79–87
Hamburger, J., 29:67–72
Hanshaw, J. B., 28:103–10
Hardwicke, J., 30:199–211
Harker, L. A., 31:361–71
Harrington, W. J., 28:299–309
Harris, J. R., 32:387–404
Hartmann, R. C., 28:187–94
Havel, R. J., 28:195–209
Hellman, S., 32:387–404
Herbert, V., 31:433–39
Hersh, E. M., 28:489–515
Hetzel, D., 30:61–79
Holland, J. F., 28:475–88
Holmsen, H., 30:119–34
Hoofnagle, J. H., 32:1–11
Hook, E. W., 28:371–91
Hornick, R. B., 30:457–72
Hoyumpa, A. M. Jr., 29:205–18
Hunder, G. G., 29:15–22
Hunt, S. A., 32:213–20
Hutchins, G. M., 31:289–301

I

Irsigler, G. B., 31:109–26

J

Jenkins, C. D., 29:543–62

K

Kadir, S., 30:41–59
Kales, A., 30:301–12
Kales, J. D., 30:301–12
Kannel, W. B., 32:427–34
Kaplan, N. M., 29:31–40
Karnovsky, M. J., 30:213–24
Katz, A. I., 32:273–89
Kaufman, H. E., 28:21–35
Kawai, C., 31:139–57
Kawamura, K., 31:139–57
Kaye, M. D., 30:375–404
Kestenbaum, D., 32:443–56
Kim, Y., 29:319–25; 31:273–88
Kim, Y. S., 29:99–116
Kinlough-Rathbone, R. L., 31:89–96
Kinney, D. K., 29:459–73
Kitabchi, A. E., 30:339–57
Kitzmiller, J. L., 30:375–404
Kjellstrand, C. M., 31:345–60
Klein, D. F., 32:583–99
Klein, P. A., 28:311–27
Klocke, F. J., 31:489–508
Knochel, J. P., 30:145–53
Koenig, R. J., 31:29–34
Korobkin, M., 30:181–88
Krantz, S. B., 29:51–66

L

Lamberti, J. J., 29:571–82
Langer, G. A., 28:13–20
Levene, M. B., 32:387–404
Levitt, M. D., 31:127–37
Lewis, E. J., 30:445–55
Libke, R. D., 32:161–74
Lichtenstein, L. M., 31:181–90
Liebowitz, M. R., 32:583–99
Lindheimer, M. D., 32:273–89
Linnemann, C. C. Jr., 28:179–85
Lipscomb, K., 28:271–89
Littman, P. S., 30:405–15
Lockwood, C. M., 31:167–79
Logue, G. L., 31:191–200
Lohmeier, T. E., 31:15–27
Long, J. M. III, 28:517–28
Longo, D. L., 32:475–90
Love, R., 28:131–41
Lyons, M. M., 28:393–409

M

Maas, J. W., 29:219–29
MacDonald, P. C., 30:473–88
MacLean, W. A. H., 30:259–68
Maddox, D. A., 28:91–102
Mahaffey, J. E., 29:327–42
Majerus, P. W., 29:41–49
Malarkey, W. B., 30:249–58
Mandel, L. P., 31:453–62
Marchesi, V. T., 29:593–603
Matsumori, A., 31:139–57
Matthysse, S., 29:459–73
Mavligit, G. M., 28:489–515
Mazzaferri, E. L., 32:73–91
McAnulty, J., 32:499–507
McClung, M. R., 31:385–404
McCluskey, R. T., 29:191–203
McDevitt, H. O., 28:425–52
McGuigan, J. E., 29:307–18
McPhaul, J. J. Jr., 28:37–42
Meares, E. M. Jr., 30:279–88
Merin, R. G., 28:75–83
Merrill, J. P., 29:343–58
Meyer, R. D., 31:219–32
Meyer-Bahlburg, H. F. L., 30:417–30
Michael, A. F., 29:319–25; 31:273–88
Miletich, J. P., 29:41–49
Miller, G. J., 31:97–108
Millman, B. S., 28:223–34
Minna, J. D., 30:359–74
Mitchel, D. H., 31:51–57
Modell, J. H., 29:1–8
Moellering, R. C., 32:559–81
Morris, L., 30:339–57
Morrison, J. A., 32:601–13
Morton, D. L., 29:231–83
Moss, A. J., 31:1–14
Moylan, J. A., 28:85–90
Mullins, C. B., 28:271–89
Murray, B. E., 32:559–81
Mustard, J. F., 31:89–96
Myers, F. L., 30:331–38

N

Nasenbeny, J., 28:179–85
Noël, L.-H., 29:67–72

O

Oates, J. A., 29:23–29
Odell, W. D., 29:379–406
Okada, D. M., 28:355–69
Oldewurtel, H. A., 28:393–409
Olley, P. M., 32:375–85
Orci, L., 28:119–30
Orne, M. T., 30:489–503
Osborn, A. G., 30:189–98

P

Packham, M. A., 31:89–96
Pak, C. Y. C., 28:111–17
Parisi, A. F., 28:159–66
Perloff, M., 28:475–88
Peschle, C., 31:303–14
Peters, D. K., 31:167–79
Peters, R. S., 32:161–74
Peterson, P. K., 32:29–43
Pierce, A. K., 28:211–21
Pietra, G. G., 31:421–31
Pitot, H. C., 30:25–39
Plotkin, S. A., 29:583–91
Plumb, V. J., 30:259–68
Portlock, C. S., 29:81–91
Portnoy, B. L., 28:167–77
Prusiner, S. B., 32:521–42
Purcell, R. H., 29:359–66

Q

Quest, J. A., 29:73–79
Quie, P. G., 29:285–306; 32:29–43

R

Rahimtoola, S., 32:499–507
Raisz, L. G., 29:23–29
Rapoport, B., 28:57–65
Raskin, N. H., 32:63–71

Regan, T. J., 28:393–409
Reiffel, J. A., 30:91–118
Renfrew, R., 31:345–60
Resnekov, L., 29:571–82
Robboy, S. J., 30:359–74
Robertson, G., 31:315–27
Robertson, J., 28:211–21
Robins, E., 28:67–73
Rocklin, R. E., 30:375–404
Rosenberg, R. D., 29:367–78
Ross, R., 30:1–15
Rossi, A., 29:149–75
Rubin, R. H., 30:225–39
Rude, R. K., 32:245–59
Russell, F. E., 31:247–59
Ryan, W. G., 28:143–52

S

Sacks, H., 30:339–57
Sasahara, A. A., 28:159–66
Sasazuki, T., 28:425–52
Scardino, P. T., 32:543–57
Schachter, J., 32:45–61
Schenker, S., 29:205–18
Schiavi, R. C., 32:509–20
Schiff, G. M., 31:441–51
Seltzer, H. S., 31:261–72
Severinghaus, J. W., 31:109–26
Seyberth, H. W., 29:23–29
Shah, P. M., 28:235–50
Shah, S. A., 29:563–70
Sharma, G. V. R. K., 28:159–66
Shearman, D. J. C., 30:61–79
Shimm, D. S., 31:191–200
Siegel, J. H., 32:175–94
Silva, J., Jr., 32:327–33
Simone, J. V., 32:207–12
Singer, F. R., 32:245–59
Sipes, J. N., 28:371–91
Skinner, D. G., 32:543–57
Slichter, S. J., 31:509–40
Smith, J. W., 31:373–83
Smith, R. T., 28:311–27
Snyder, W. H. III, 28:111–17
Soldatos, C. R., 30:301–12
Soll, A. H., 29:495–507
Staub, N. C., 32:291–312
Stein, E. A., 32:601–13
Stephen, C. R., 28:153–57
Stern, W. H., 32:101–6
Stewart, A. F., 32:457–73
Stinson, E. B., 32:213–20
Stossel, T. P., 29:427–57
Stropes, L., 31:315–27
Stryckmans, P. A., 31:59–66
Sweeney, D. R., 29:219–29

T

Tancini, G., 29:149–75
Tannen, R. L., 31:35–49
Tarazi, R. C., 29:485–93
Theml, H., 28:131–41
Thom, T. J., 32:427–34
Thompson, R. G., 31:453–62
Thompson, R. L., 28:371–91
Tolkoff-Rubin, N. E., 30:225–39
Tomasi, T. B. Jr., 28:453–65
Tow, D. E., 28:159–66

U

Unger, R. H., 28:119–30

V

Verhoef, J., 32:107–22
Verbrugh, H. A., 32:107–22
Voorhees, J. J., 28:467–73

W

Waldo, A. L., 30:259–68
Wall, R. T., 31:361–71
Walzer, S., 29:563–70
Ward, L. E., 29:15–22
Warren, J. W., 30:457–72
Wechsler, A. S., 28:341–48
Weisfeldt, M. L., 32:435–42
Weiss, H. J., 30:119–34
Wells, J. L. Jr., 30:259–68
Whitley, R., 32:335–40
Wiktor, T., 29:583–91
Wilfert, C. M., 29:129–36
Williams, R. C. Jr., 32:13–28
Wilson, D. R., 28:329–39
Winkelmann, R. K., 28:251–59
Wisneski, J. A., 29:475–83
Wolfe, M. S., 31:373–83
Wolfsen, A. R., 29:379–406

Y

Yalow, R. S., 30:241–48
Young, D. B., 31:15–27
Young, R., 30:339–57
Young, R. C., 32:475–90

Z

Zerbe, R., 31:315–27
Ziegler, M. G., 31:239–45
Zipes, D. P., 29:417–26
Zusman, R. M., 32:359–74

CHAPTER TITLES, VOLUMES 28–32

ANESTHESIOLOGY		
Effects of Anesthetic Drugs on Myocardial Performance in Man	R. G. Merin	28:75–83
Malignant Hyperpyrexia	C. R. Stephen	28:153–57
Acupuncture: Context and Critique	B. S. Millman	28:223–34
CARDIOVASCULAR SYSTEM		
The Detection of Asymptomatic Coronary Artery Disease	V. F. Froelicher Jr.	28:1–12
Ionic Basis of Myocardial Contractility	G. A. Langer	28:13–20
Hypertrophic Cardiomyopathy	P. M. Shah	28:235–50
Development of Coronary Collateral Circulation	A. S. Wechsler	28:341–48
Prophylaxis of Infective Endocarditis: A Reevaluation	J. N. Sipes, R. L. Thompson, E. W. Hook	28:371–91
The Role of Ethanol in Cardiac Disease	T. J. Regan, P. O. Ettinger, B. Haider, S. S. Ahmed, H. A. Oldewurtel, M. M. Lyons	28:393–409
Neural Actions of Digitalis	R. A. Gillis, J. A. Quest	29:73–79
Slow-Channel Depolarization: Mechanism and Control of Arrhythmias	J. C. Bailey, V. Elharrar, D. P. Zipes	29:417–26
Left Ventricular Stiffness	J. A. Wisneski, J. D. Bristow	29:475–83
Cardiogenic Hypertension	H. P. Dustan, R. C. Tarazi	29:485–93
Behavioral Risk Factors in Coronary Artery Disease	J. D. Jenkins	29:543–62
Cardiac Assist Devices	J. J. Lamberti, L. Resnekov	29:571–82
The Arterial Wall and Atherosclerosis	R. Ross	30:1–15
Validity of Therapy for Mild Hypertension	E. D. Freis	30:81–89
Sick Sinus Syndrome	J. T. Bigger Jr., J. A. Reiffel	30:91–118
The Prolonged QT Interval	J. A. Abildskov	30:171–79
Studies of Atrial Flutter Following Open Heart Surgery	A. L. Waldo, J. L. Wells, Jr., V. J. Plumb, T. B. Cooper, W. A. H. MacLean	30:259–68
Regression of Atherosclerosis	D. W. Crawford, D. H. Blankenhorn	30:289–300
Prediction and Prevention of Sudden Cardiac Death	A. J. Moss	31:1–14
Salt Balance and Long-Term Blood Pressure Control	A. C. Guyton, T. G. Coleman, D. B. Young, T. E. Lohmeier, J. W. DeClue	31:15–27
High Density Lipoproteins and Atherosclerosis	G. J. Miller	31:97–108
Myocardial Biopsy	C. Kawai, A. Matsumori, K. Kawamura	31:139–57
Noninvasive Evaluation of the Carotid Bruit	R. W. Barnes	31:201–18
Postural Hypotension	M. G. Ziegler	31:239–45
Pathological Changes in Aortocoronary Bypass Grafts	G. M. Hutchins	31:289–301
Pericarditis and Renal Failure	R. Renfrew, T. J. Buselmeier, C. M. Kjellstrand	31:345–60

Primary Pulmonary Hypertension	A. P. Fishman, G. G. Pietra	31:421–31
Community Cardiopulmonary Resuscitation	L. A. Cobb, A. P. Hallstrom, R. G. Thompson, L. P. Mandel, M. K. Copass	31:453–62
Control of Coronary Blood Flow	F. J. Klocke, A. K. Ellis	31:489–508
Prostaglandins and the Ductus Arteriosus	P. M. Olley, F. Coceani	32:375–85
Downward Trend in Cardiovascular Mortality	T. J. Thom, W. B. Kannel	32:427–34
Physiology of Cardiopulmonary Resuscitation	M. L. Weisfeldt, N. Chandra	32:435–42
Prognosis in Bundle Branch Block	J. McAnulty, S. Rahimtoola	32:499–507
Coronary Risk Factors in the Young	E. A. Stein, C. J. Glueck, J. A. Morrison	32:601–13

CLINICAL PHARMACOLOGY

The Disposition and Effects of Sedatives and Analgesics in Liver Disease	A. M. Hoyumpa, Jr., R. A. Branch, S. Schenker	29:205–18
Clinical Pharmacology of Acute Allergic Disorders	J. C. Foreman, L. M. Lichtenstein	31:181–90
Cephalosporins	B. E. Murray, R. C. Moellering	32:559–81

DERMATOLOGY

Current Problems in Mycosis Fungoides and Sezary Syndrome	R. K. Winkelmann, W. A. Caro	28:251–69
Pathophysiology of Psoriasis	J. J. Voorhees	28:467–73
Chronic Mucocutaneous Candidiasis	J. M. Dwyer	32:491–97

ENDOCRINOLOGY AND METABOLISM

Impaired Peripheral Conversion of Thyroxine to Triiodthyronine	R. R. Cavalieri, B. Rapoport	28:57–65
Diagnosis of Hyperparathyroidism	H. C. Bone, III, W. H. Snyder, C. Y. C. Pak	28:111–17
The Role of Glucagon in the Endogenous Hyperglycemia of Diabetes Mellitus	R. H. Unger, L. Orci	28:119–30
Paget's Disease of Bone	W. G. Ryan	28:143–52
Classification of the Hyperlipidemias	R. J. Havel	28:195–209
Prostaglandins and Hypercalcemic States	H. W. Seyberth, L. G. Raisz, J. A. Oates	29:23–29
Clinical Hypophosphatemia	F. Fitzgerald	29:177–89
Osteomalacia and Disorders of Vitamin D Metabolism	J. F. Habener, J. E. Mahaffey	29:327–42
Humoral Syndromes Associated with Cancer	W. D. Odell, A. R. Wolfsen	29:379–406
Medical Treatment of Acromegaly	I. D. Goldfine	29:407–15
Control of the Onset of Puberty	R. M. Boyar	29:509–20
The Syndrome of Hyporeninemic Hypoaldosteronism	J. P. Knochel	30:145–53
Dietary Management of Diabetes Mellitus	J. W. Ensinck, E. L. Bierman	30:155–70
Big ACTH and Bronchogenic Carcinoma	R. S. Yalow	30:241–48
Prolactin and the Diagnosis of Pituitary Tumors	W. B. Malarkey	30:249–58
Diagnosis and Medical Management of Insulinomas	S. S. Fajans, J. C. Floyd, Jr.	30:313–29
Diabetic Ketoacidosis: Reappraisal of Therapeutic Approach	A. E. Kitabchi, R. Young, H. Sacks, L. Morris	30:339–57
Hemoglobin A_{1c} and Diabetes Mellitus	R. J. Koenig, A. Cerami	31:29–34
Efficacy and Safety of Oral Hypoglycemic Agents	H. S. Seltzer	31:261–72
Treatment of Hyperthyroidism	M. R. McClung, M. A. Greer	31:385–404
Bartter's Syndrome	J. R. Gill	31:405–19
Papillary and Follicular Thyroid Cancer: A Selective Approach to Diagnosis and Treatment	E. L. Mazzaferri	32:73–91
The Problem of Obesity	F. T. Fitzgerald	32:221–31

Disorders of Glutamate Metabolism and Neurological Dysfunction	S. B. Prusiner	32:521–42
GASTROINTESTINAL SYSTEM		
Viral Agents and Diarrheal Illness	H. L. DuPont, L. Portnoy, R. H. Conklin	28:167–77
Applications and Hazards of Total Intravenous Hyperalimentation	S. J. Dudrick, J. M. Long, III	28:517–28
Digestion and Absorption of Protein	H. J. Freeman, Y. S. Kim	29:99–116
Gastrointestinal Hormones	J. E. McGuigan	29:307–18
Non-A, Non-B Hepatitis	S. M. Feinstone, R. H. Purcell	29:359–66
Cellular Mechanisms in Acid Secretion	A. H. Soll, M. I. Grossman	29:495–507
Angiographic Management of Gastrointestinal Bleeding	S. Kadir, C. A. Athanasoulis	30:41–59
The Medical Management of Peptic Ulcer	D. J. C. Shearman, D. Hetzel	30:61–79
Computed Tomography of the Hepatobiliary System	M. Korobkin, H. I. Goldberg	30:181–88
Amanita Mushroom Poisoning	D. H. Mitchel	31:51–57
The Medical Treatment of Gallstones	I. A. D. Bouchier	31:59–77
Intestinal Absorption and Malabsorption of Folates	C. H. Halsted	31:79–87
Flatulence	M. D. Levitt, J. H. Bond	31:127–37
Folate-Binding Proteins	N. Colman, V. Herbert	31:433–39
Serological Markers of Hepatitis B Virus Infection	J. H. Hoofnagle	32:1–11
Pathogenesis of Acute Bacterial Diarrheal Disorders	R. A. Giannella	32:341–57
Pathogenesis, Diagnosis, and Management of Reflux Esophagitis	D. Kestenbaum, J. Behar	32:443–56
HEMATOLOGY		
Factors in the Pathomechanism of Chronic Lymphocytic Leukemia	H. Theml, R. Love, H. Begemann	28:131–41
Paroxysmal Nocturnal Hemoglobinuria (PNH) as a Clonal Disorder	R. C. Hartmann, A. B. Arnold	28:187–94
Cold Agglutinins and Cold-Agglutinin Disease	M. M. Frank, J. P. Atkinson, J. Gadek	28:291–98
Treatment of Idiopathic Thrombocytopenic Purpura (ITP)	Y. S. Ahn, W. J. Harrington	28:299–309
Relationships Between Platelets and Coagulation Factors in Hemostasis	P. W. Majerus, J. P. Miletich	29:41–49
Erythropoietin and the Control of Red Cell Production	S. E. Graber, S. B. Krantz	29:51–66
Heparin, Antithrombin, and Abnormal Clotting	R. D. Rosenberg	29:367–78
Contractile Proteins in Cell Structure and Function	T. P. Stossel	29:427–57
Hairy Cell Leukemia	R. C. Braylan, J. S. Burke	30:17–24
Secretable Storage Pools in Platelets	H. Holmsen, H. J. Weiss	30:119–34
Disseminated Intravascular Coagulation: A Reappraisal	R. W. Colman, S. J. Robboy, J. D. Minna	30:359–74
Treatment of Plasma Cell Myeloma	D. E. Bergsagel	30:431–43
Prostaglandins and Platelets	J. F. Mustard, R. L. Kinlough-Rathbone, M. A. Packham	31:89–96
Treatment of Chronic Myeloid Leukemia	P. A. Stryckmans	31:159–66
Autoimmune Granulocytopenia	G. L. Logue, D. S. Shimm	31:191–200
Erythropoiesis	C. Peschle	31:303–14
The Endothelium and Thrombosis	R. T. Wall, L. A. Harker	31:361–71
Folate-Binding Proteins	N. Colman, V. Herbert	31:433–39

Controversies in Platelet Transfusion Therapy	S. J. Slichter	31:509–40
Anemia of Prematurity	P. R. Dallman	32:143–60
Treatment of Sickle Cell Anemia	S. Charache	32:195–206
Outlook for Acute Lymphocytic Leukemia in Children in 1982	J. V. Simone	32:207–12
Amyloidosis	P. D. Gorevic, E. C. Franklin	32:261–71

IMMUNOLOGY

The Association of Genes in the Major Histocompatibility Complex and Disease Susceptibility	T. Sasazuki, F. C. Grumet, H. O. McDevitt	28:425–52
Control of the Immune System by Inhibitor and Inducer T Lymphocytes	H. Cantor	30:269–77
Immunobiology of the Maternal-Fetal Relationship	R. E. Rocklin, J. L. Kitzmiller, M. D. Kaye	30:375–404
Immune Complexes in Human Diseases	R. C. Williams, Jr.	32:13–28

INFECTIOUS DISEASES

Cytomegalovirus (CMV) in the Compromised Host(s)	R. F. Betts, J. B. Hanshaw	28:103–10
Pertussis in the Adult	C. C. Linnemann Jr., J. Nasenbeny	28:179–85
The Emergence of Group B Streptococci in Infections of the Newborn Infant	B. F. Anthony, D. M. Okada	28:355–69
The Current Status of BCG Immunization Against Tuberculosis	T. C. Eickhoff	28:411–23
E. Coli Meningitis: K_1 Antigen and Virulence	C. M. Wilfert	29:129–36
Otitis Media: The Spectrum of Middle Ear Inflammation	G. S. Giebink, P. G. Quie	29:285–306
Rabies Vaccination	S. A. Plotkin, T. Wiktor	29:583–91
Immunization Against Typhoid Fever	J. W. Warren, R. B. Hornick	30:457–72
Legionnaires' Disease	R. D. Meyer, S. M. Finegold	31:219–32
Pulmonary Function During and After Common Respiratory Diseases	W. J. Hall, R. G. Douglas, Jr.	31:233–38
Giardiasis	J. W. Smith, M. S. Wolfe	31:373–83
Active Immunization for Adults	G. M. Schiff	31:441–51
Infant Botulism	S. S. Arnon	31:541–60
Serological Markers of Hepatitis B Virus Infection	J. H. Hoofnagle	32:1–11
Bacterial Surface Components and the Pathogenesis of Infectious Diseases	P. K. Peterson, P. G. Quie	32:29–43
Chlamydial Infections	J. Schachter, M. Grossman	32:45–61
Host Determinants in Staphylococcal Disease	J. Verhoef, H. A. Verbrugh	32:107–22
Amebic Liver Abscess	R. S. Peters, N. Gitlin, R. D. Libke	32:161–74
Relations Between Circulatory and Metabolic Changes in Sepsis	J. H. Siegel	32:175–94
The Killing of Pathogens by Phagocytes	T. G. Gabig, B. M. Babior	32:313–26
Clostridia and Antimicrobial Enterocolitis	J. Silva, Jr., R. Fekety	32:327–33
Diagnosis and Treatment of Herpes Simplex Encephalitis	R. Whitley	32:335–40
Pathogenesis, Diagnosis, and Management of Reflux Esophagitis	D. Kestenbaum, J. Behar	32:443–56

KIDNEY AND URINARY SYSTEM

IgA-Associated Glomerulonephritis	J. J. McPhaul, Jr.	28:37–42
Glomerular Filtration of Fluid and Macromolecules: The Renal Response to Injury	D. A. Maddox, B. M. Brenner	28:91–102

Renal Function During and Following Obstruction	D. R. Wilson	28:329–39
Urate Excretion	G. M. Fanelli, Jr.	28:349–54
Recurrent Glomerulonephritis After Renal Transplantation	J. Hamburger, J. Crosnier, L.-H. Noël	29:67–72
Dialysis Encephalopathy Syndrome	A. C. Alfrey	29:93–98
Immunological Aspects of Renal Tubular and Interstitial Diseases	R. T. McCluskey, R. B. Colvin	29:191–203
Chronic Bacteremia and Nephritis	Y. Kim, A. F. Michael	29:319–25
Dialysis Versus Transplantation in the Treatment of End-Stage Renal Disease	J. P. Merrill	29:343–58
Renal Tubular Acidosis	F. J. Gennari, J. J. Cohen	29:521–41
Low Molecular Weight Proteinuria	C. L. Hall, J. Hardwicke	30:199–211
The Ultrastructure of Glomerular Filtration	M. J. Karnovsky	30:213–24
Localization and Antibiotic Management of Urinary Tract Infection	L. S. T. Fang, N. E. Tolkoff-Rubin, R. H. Rubin	30:225–39
Prostatitis	E. M. Meares, Jr.	30:279–88
Patterns of Circulating Complement in Renal Diseases	E. J. Lewis	30:445–55
Salt Balance and Long-Term Blood Pressure Control	A. C. Guyton, T. G. Coleman, D. B. Young, T. E. Lohmeier, J. W. DeClue	31:15–27
Control of Acid Excretion by the Kidney	R. L. Tannen	31:35–49
Plasma Exchange in Glomerulonephritis and Related Vasculitides	C. M. Lockwood, D. K. Peters	31:167–79
Idiopathic Membranoproliferative Glomerulonephritis	Y. Kim A. F. Michael	31:273–88
Vasopressin Function in the Syndrome of Inappropriate Antidiuresis	R. Zerbe, L. Stropes, G. Robertson	31:315–27
Pericarditis and Renal Failure	R. Renfrew, T. J. Buselmeier, C. M Kjellstrand	31:345–60
The Natural History and Response to Therapy of Lupus Nephritis	E. M. Ginzler, A. J. Bollet, E. A. Friedman	31:463–87
Magnesium Deficiency and Excess	R. K. Rude, F. R. Singer	32:245–59
Pathophysiology of Preeclampsia	M. D. Lindheimer, A. I. Katz	32:273–89
Prostaglandins and Water Excretion	R. M. Zusman	32:359–74
The Regulation of Renal Calcium Excretion: An Approach to Hypercalciuria	A. F. Stewart, A. E. Broadus	32:457–73

MISCELLANEOUS

Confidentiality in Medical Practice	M. Grossman	28:43–55
Informed Consent	G. J. Annas	29:9–14
The XYY Genotype	S. Walzer, P. S. Gerald, S. A. Shah	29:563–70
Recent Membrane Research and Its Implications for Clinical Medicine	V. T. Marchesi	29:593–603
The Efficacy of Biofeedback Therapy	M. T. Orne	30:489–503
Snake Venom Poisoning in the United States	F. E. Russell	31:247–59
Retinal Vascular Occlusion	W. H. Stern, D. B. Archer	32:101–6

NEOPLASIA

The Role of Oncogenic Viruses in Neoplasia	P. A. Klein, R. T. Smith	28:311–27
Structure and Function of Alpha-Fetoprotein	T. B. Tomasi, Jr.	28:453–65
Surgical Adjuvant Chemotherapy	M. Perloff, J. F. Holland	28:475–88
BCG as Adjuvant Immunotherapy for Neoplasia	E. M. Hersh, J. U. Gutterman, G. M. Mavligit	28:489–515
The Non-Hodgkin's Lymphomas: Current Concepts and Management	C. S. Portlock, E. Glatstein	29:81–91

Chemotherapy in Prevention of the Recurrence of Resectable Cancer	G. Bonadonna, G. Tancini, A. Rossi, M. Gasparini	29:149–75
Immunotherapy for Malignant Disease	J. E. Goodnight, Jr., D. L. Morton	29:231–83
Biological and Enzymatic Events in Chemical Carcinogenesis	H. C. Pitot	30:25–39
Clonal Origin of Human Tumors	P. J. Fialkow	30:135–43
Growth Considerations in the Radiation Therapy of Children with Cancer	P. S. Littman, G. J. D'Angio	30:405–15
Primary Radiation Therapy for Breast Cancer	J. R. Harris, M. B. Levene, S. Hellman	32:387–404
The Natural History and Treatment of Ovarian Cancer	D. L. Longo, R. C. Young	32:475–90
Testicular Cancer	D. G. Skinner, P. T. Scardino, J. R. Daniels	32:543–57

NERVOUS SYSTEM

Computed Tomography in Neurologic Diagnosis	A. G. Osborn	30:189–98
Prenatal Sex Hormones and the Developing Brain: Effects on Psychosexual Differentiation and Cognitive Function	A. A. Ehrhardt, H. F. L. Meyer-Bahlburg	30:417–30
Chemical Headaches	N. H. Raskin	32:63–71
Dietary Precursors and Brain Neurotransmitter Formation	J. D. Fernstrom	32:413–25
Diagnosis and Treatment of Herpes Simplex Encephalitis	R. Whitley	32:335–40
Disorders of Glutamate Metabolism and Neurological Dysfunction	S. B. Prusiner	32:521–42

PSYCHIATRY

New Concepts in the Diagnosis of Psychiatric Disorders	E. Robins	28:67–73
Anorexia Nervosa: Recent Investigations	K. A. Halmi	29:137–48
Specificity of Depressive Diseases	D. R. Sweeney, J. W. Maas	29:219–29
Genetic Transmission of Schizophrenia	D. K. Kinney, S. Matthysse	29:459–73
Management of Insomnia	C. R. Soldatos, A. Kales, J. D. Kales	30:301–12
Genetic Component of Alcoholism	D. W. Goodwin	32:93–99
Neuroendocrinological Effects of Alcohol	T. J. Cicero	32:123–42
Convulsive and Drug Therapies of Depression	M. Fink	32:405–12
Male Erectile Disorders	R. C. Schiavi	32:509–20
Differential Diagnosis and Treatment of Panic Attacks and Phobic States	M. R. Liebowitz, D. F. Klein	32:583–99

REPRODUCTIVE SYSTEM

Cardiovascular Complications of Oral Contraceptives	N. M. Kaplan	29:31–40
The Prenatal Diagnosis of Genetic Disorders	C. J. Epstein, M. S. Golbus	29:117–28
Endocrinology of the Placenta	R. B. Everett, P. C. MacDonald	30:473–88
Effectiveness and Risks of Contraception	W. Droegemueller, R. Bressler	31:329–43

RESPIRATORY SYSTEM

Diagnosis and Treatment of Fat Embolism	J. A. Moylan, M. A. Evenson	28:85–90
Diagnosis of Pulmonary Embolism	G. V. R. K. Sharma, D. E. Tow, A. F. Parisi, A. A. Sasahara	28:159–66
Biology of Drowning	J. H. Modell	29:1–8
Clinical Problems of Ventilatory Control	G. B. Irsigler, J. W. Severinghaus	31:109–26
Pulmonary Function During and After Common Respiratory Diseases	W. J. Hall, R. G. Douglas, Jr.	31:233–38
Primary Pulmonary Hypertension	A. P. Fishman, G. G. Pietra	31:421–31

Pulmonary Edema Due To Increased Microvascular Permeability	N. C. Staub	32:291–312
RHEUMATOLOGY		
Polymyalgia Rheumatica and Giant Cell Arteritis	R. E. Ettlinger, G. C. Hunder, L. E. Ward	29:15–22
SURGICAL TOPICS		
Corneal Transplantation	S. L. Forstot, H. E. Kaufman	28:21–35
Pulmonary Complications of General Surgery	A. K. Pierce, J. Robertson	28:211–21
Critique of Coronary Artery Bypass Surgery	C. B. Mullins, K. Lipscomb	28:271–89
Vitrectomy Surgery for Diabetic Retinopathy	G. H. Bresnick, F. L. Myers	30:331–38
Amebic Liver Abscess	R. S. Peters, N. Gitlin, R. D. Libke	32:161–74
Relations Between Circulatory and Metabolic Changes in Sepsis	J. H. Siegel	32:175–94
Cardiac Transplantation	S. A. Hunt, E. B. Stinson	32:213–20
Total Parenteral Nutrition in the Management of the Cancer Patient	M. F. Brennan	32:233–43